

**Metodologia para a construção e análise de *corpora* de especialidade para fins  
específicos: o caso da Segurança Automóvel**

**Américo José Lopes Redondo**

**Dissertação de Mestrado em Terminologia e Gestão de Informação de  
Especialidade**

**Setembro, 2016**

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de  
Mestre em Terminologia e Gestão de Informação de Especialidade

Realizada sob a orientação científica da

Prof.<sup>a</sup> Doutora Rute Costa

## DECLARAÇÕES

Declaro que esta Dissertação é o resultado da minha investigação pessoal e independente. O seu conteúdo é original e todas as fontes consultadas estão devidamente mencionadas no texto, nas notas e na bibliografia.

O candidato,

---

Lisboa, .... de ..... de .....

Declaro que esta Dissertação se encontra em condições de ser apreciado pelo júri a designar.

O(A) orientador(a),

---

Lisboa, .... de ..... de .....

## **AGRADECIMENTOS**

A redação de uma dissertação é uma tarefa difícil e suportar e superar esta dificuldade não foi um esforço solitário.

Foi a partir das conversa com colegas e professores que se começou a formar esta aventura. As suas opiniões e sugestões contribuíram para o nascimento da ideia por detrás da dissertação, muitas vezes em ambiente de conversa casual e descontraída.

A parte mais difícil calhou à Prof.<sup>a</sup> Doutora Rute Costa, que no seu papel de orientadora teve sempre uma conduta exigente, mas sempre com a intenção de avivar a minha motivação, foi paciente e mostrou-se sempre disponível para ajudar.

A convivência com amigos e familiares foi importante. O tempo passado com estes possibilitou momentos de descontração e distração que permitiram recarregar baterias para poder enfrentar as teclas com mais vontade.

Pela vossa amizade, a todos deixo o meu sentido muito obrigado.

# **Metodologia para a construção e análise de *corpora* de especialidade para fins específicos: o caso da Segurança Automóvel**

Américo José Lopes Redondo

## **RESUMO**

**PALAVRAS-CHAVE:** terminologia; linguística de *corpus*; segurança automóvel; *corpora*; termo; metodologia; unidade lexical; veículos

A realização deste trabalho decorre do pressuposto de que é possível obter um recurso viável para ser submetido a análise linguística a partir de informação obtida na Internet. Reconhecendo a importância da internet como um repositório de informação e conhecimento, entendemos que existe lugar ao seu aproveitamento para proceder à construção de *corpora* de especialidade. Pretende-se construir uma metodologia, fundamentada em aptidões adquiridas a partir do trabalho e aprendizagem operados na investigação e construção de *corpora* com cariz terminológico. Esta abordagem pretende focar-se na componente prática, embora seja necessário recorrer a pressupostos teóricos das áreas da linguística de *corpora* e da Terminologia.

Demonstramos que é possível obter uma metodologia que aproveita as capacidades de uma ferramenta de tratamento automático da língua para selecionar unidades lexicais que têm grande possibilidade de se estabelecerem como termos de um domínio, reconhecendo a necessidade de recorrer à competência de especialistas para uma subsequente confirmação. Esta metodologia permite que, alunos, professores e investigadores construam *corpora* de especialidade e procedam à sua análise de uma forma autónoma. Consideramos que a metodologia obtida pode servir para complementar a falta de recursos de estudo na área de ensino de línguas de especialidade.

# **Specialized *corpora* construction and analysis methodology for specific purposes: the case of automobile safety**

Américo José Lopes Redondo

## **ABSTRACT**

**KEYWORDS:** Terminology; *corpus* linguistics; automobile safety; *corpora*; term; methodology; lexical unit; vehicles

This work results from the assumption that it is possible to produce a viable resource to be subjected to linguistic analysis from information obtained from the Internet. Recognizing the importance of the Internet as a repository of information and knowledge, we realize the advantage in using it in the construction of specialized *corpora*. The goal is to build a methodology, based on the skills acquired from the work and learning carried out while investigating and constructing *corpora* of terminological nature. This approach aims to focus on the practical component, although it is necessary to use theoretical assumptions concerning the subjects of *corpus* linguistics and terminology.

We demonstrate that, it is possible to obtain a methodology that takes the potential of an automatic language processing tool to select lexical units which have great potential to establish themselves as terms of a domain, recognizing the need to call upon the competence of specialists for subsequent confirmation. This methodology allows students, teachers and researchers to construct specialized *corpora* and proceed to its analysis in an independent fashion. We consider that the methodology obtained can be used to supplement the lack of study resources in the teaching of specialized languages

## ÍNDICE

Introdução .....	1
Capítulo 1: A segurança automóvel .....	3
1.1. Consulta de dicionários online .....	3
1.2. Consulta de sítios .....	6
1.2.1. Hiperligação 1 .....	7
1.2.2. Hiperligação 2 .....	9
1.2.3. Hiperligação 3 .....	10
1.3. Áreas de atividade das entidades: CE, IMT, ANSR .....	11
1.4. Descrição dos conceitos de base do domínio .....	12
1.4.1. Mapas conceptuais.....	16
Capítulo 2: Linguística de <i>corpus</i> .....	18
2.1 <i>Corpora</i> e ensino da língua .....	19
2.2 <i>Corpora</i> e Terminologia .....	19
Capítulo 3: Constituição do <i>corpus</i> .....	21
3.1. Seleção dos textos .....	22
3.2. Problemas encontrados .....	24
Capítulo 4: Tratamento semiautomático do <i>corpus</i> : ferramentas de análise de <i>corpora</i> .....	26
4.1. O AntConc: principais características .....	27
Capítulo 5: Determinação do campo semântico do domínio .....	34
5.1. Seleção de termos.....	37
5.2. Análise de uma forma pré-selecionada - [entende-se] .....	48
5.2.1. Análise da combinação das formas [entende-se um].....	49

5.2.2. Análise da combinação das formas [entende-se por] .....	55
5.2.3. Observações.....	57
Conclusão .....	58
Bibliografia .....	60
Índice de Figuras .....	65
Índice de Tabelas .....	67



## CONVENÇÕES GRÁFICAS

No corpo do texto são usadas as seguintes convenções gráficas:

- a. < > assinala um conceito;
- b. ' ' assinala um termo;
- c. // assinala uma unidade lexical;
- d. [ ] assinala uma forma

## Introdução

É nossa intenção constituir uma metodologia que possa ser aplicada na criação de recursos terminológicos que possam servir vários fins, entre os quais o de apoiar o ensino da língua em contexto de especialidade. Entenda-se, contudo, que não se pretende com este trabalho intervir em aspetos didáticos. A nossa proposta pretende propor uma metodologia que se constitua como um recurso a considerar, por professores e alunos, como complemento aos planos estabelecidos institucionalmente, partindo do nosso conhecimento das práticas de análise e investigação de *corpora* em Terminologia.

A abordagem neste estudo é a de trabalhar *corpora* compostos por textos de especialidade escritos em Português Europeu. A metodologia apresentada neste trabalho tem por finalidade demonstrar, independentemente da língua de trabalho, que é possível recorrer a princípios da linguística de *corpora* de especialidade para, com os resultados obtidos, construir recursos terminológicos que possam auxiliar o ensino-aprendizagem de uma língua de especialidade, seja ela materna, segunda ou estrangeira.

Para a construção da metodologia, procedemos à descrição das várias etapas a seguir:

- a. Descrição do domínio em análise;
- b. Seleção dos textos de especialidade;
- c. Constituição do *corpus*;
- d. Tratamento semiautomático do *corpus*;
- e. Análise dos dados;

Na escolha do domínio, optámos pela Segurança Automóvel<sup>1</sup> que tem um alto nível de diferenciação, uma vez que escolhemos textos que foram produzidos por especialistas. Tendo que constituir um *corpus* de especialidade, é importante determinar o domínio de especialidade, para poder compilar textos que possam integrar o *corpus*.

Como se pode verificar, esta dissertação é predominantemente metodológica. Contudo, para cumprir com o objetivo estabelecido torna-se necessário refletir sobre alguns aspetos teóricos. Com o propósito de determinar, por exemplo, um campo semântico a partir do qual se possa desenvolver uma metodologia de extração de termos, será necessário distinguir entre forma, unidade lexical e termo. Além disso, importa ainda dar conta de

---

<sup>1</sup> Sempre que ocorrer a combinação Segurança Automóvel com letras iniciais maiúsculas estaremos a referir-nos ao domínio.

algumas características referentes à disciplina da linguística de *corpus*, destacando as suas possibilidades de aplicação no ensino de línguas e em Terminologia

Por ser fundamental perceber o processo de compilação do *corpus*, descrevemos o processo referindo os principais problemas encontrados, passando depois à enumeração de alguns dos principais programas de análise de *corpora* existentes e a uma descrição do programa utilizado neste trabalho, o AntConc.

A metodologia que segue os passos atrás enunciados parte da análise de dois tipos de dados mais específicos:

- a. unidades lexicais candidatas a termos
- b. formas que estabelecem relações entre os candidatos a termos passíveis de identificar estruturas próprias de contextos definitórios

De um ponto de vista teórico, a metodologia decorre fortemente de fundamentos da terminologia e da Linguística de *Corpora*. Do ponto de vista prático, a metodologia decorre do uso de ferramentas de tratamento semiautomático da língua.

Na conclusão, damos conta do caminho percorrido durante o trabalho, sem esquecer as possibilidades a explorar no futuro, em função dos problemas encontrados e dos objetivos alcançados.

## **1. A segurança automóvel**

O produto automóvel tem uma presença a nível global, sendo que a sua fabricação tem de cumprir regras específicas e normalizadas, criadas por entidades reguladoras, nacionais e internacionais com responsabilidades em áreas da sociedade como o direito, a saúde pública e a economia.

A preocupação com a segurança está presente em muitas atividades e a segurança dos veículos automóveis preocupa os fabricantes de automóveis uma vez que contribui para a qualidade do produto e para o aumento da confiança dos utilizadores.

A Segurança Automóvel é uma área de especialização regulada por entidades com competências ao nível da criação de legislação com o objetivo de estabelecer normas a nível internacional, no sentido de regulamentar a produção, a comercialização e a utilização dos automóveis.

A Segurança Automóvel é um domínio em constante mudança. O desenvolvimento de novas tecnologias e o aparecimento de novos materiais pressupõe que a sua fiscalização e regulamentação sejam feitas por entidades com autoridade para tal, que assumam a responsabilidade de melhorar a segurança automóvel através da implementação de legislação específica, contribuindo assim para a confiança dos utilizadores dos automóveis.

### **1.1. Consulta de dicionários online**

Com o propósito de obter mais informação acerca do domínio da Segurança Automóvel, optámos por pesquisar a forma [segurança automóvel] em dois dicionários online de língua corrente, a saber o da Priberam<sup>2</sup> e o da Infopédia<sup>3</sup>.

O recurso ao dicionário de língua corrente justifica-se pelo facto de ser uma fonte de informação cuja consulta permite preencher falhas no nosso conhecimento acerca de uma ou várias palavras.

Através da pesquisa da forma [segurança automóvel] no dicionário da Priberam não obtivemos resposta direta. Neste dicionário [segurança automóvel] não corresponde a uma entrada lexicográfica. No entanto, reenviou-nos para quatro termos relacionados com a forma

---

<sup>2</sup> <http://www.priberam.pt/dlpo/seguran%C3%A7a%20autom%C3%B3vel>

<sup>3</sup> <http://www.infopedia.pt/dicionarios/lingua-portuguesa/seguran%C3%A7a%20autom%C3%B3vel>

[automóvel], a saber: ‘freio’, ‘travão’, ‘caravana’ e ‘airbag’, tal como podemos ver na imagem abaixo:

**Figura 1 – Dicionário Priberam**



No caso do dicionário da Infopédia, obtivemos a seguinte informação:

**Figura 2 – Dicionário Infopédia**



Da pesquisa realizada obtivemos resultados separados para as duas formas do termo ‘segurança automóvel’, com [automóvel] em primeiro lugar, logo seguido de [segurança]. Estas duas formas correspondem aos dois termos simples que formam o termo complexo ‘segurança automóvel’. Cada uma das definições dos termos simples parece remeter-nos para

o domínio em análise. Esta foi esta a razão que nos levou a pesquisá-los de forma independente.

Para os termos ‘automóvel’ e ‘segurança’, obtivemos os seguintes resultados:

**Tabela 1 -- Definições para ‘automóvel’ e ‘segurança’**

<b>Termo</b>	<b>Infopédia <sup>4</sup></b>	<b>Priberam <sup>5</sup></b>
<b>automóvel</b>	adjetivo 1. (veículo) com capacidade de locomoção autónoma nome masculino 2.veículo de pelo menos quatro rodas, com motor próprio (acionado geralmente a gasolina, gasóleo ou gás), usado no transporte de passageiros e de mercadorias	substantivo masculino 1. Veículo que se move por meio automático. adjectivo de dois géneros 2. Que se move automaticamente
<b>segurança</b>	nome feminino 1. ato ou efeito de segurar 2. confiança 3. certificação 4. tranquilidade de espírito 5. caução; garantia 6. firmeza; certeza; convicção 7. amparo 8. afirmação 9. corpo de vigilância e proteção contra possíveis atentados ou ataques a uma instituição ou personalidade	substantivo feminino 1. Acto ou efeito de segurar. 2. Qualidade do que é ou está seguro. ≠ INSEGURANÇA 3. Conjunto das acções e dos recursos utilizados para proteger algo ou alguém. 4. O que serve para diminuir os riscos ou os perigos. = GARANTIA 5. Aquilo que serve de base ou que dá estabilidade ou apoio. = AMPARO, ESTEIO 6. Sentimento de força interior ou de crença em si mesmo. = CERTEZA, CONFIANÇA, FIRMEZA ≠ INSEGURANÇA 7. Afoiteza, ousadia. 8. Força ou convicção nos movimentos ou nas acções. 9. Certeza demonstrada. = EVIDÊNCIA 10. Caução. substantivo de dois géneros 11. Pessoa cuja actividade profissional consiste em proteger pessoas, instalações ou bens, ou em controlar o acesso de pessoas a determinado local.

Por a Infopédia e a Priberam serem dicionários de língua corrente, não fornecem informação suficiente para podermos afirmar que o conceito designado pelo termo ‘segurança automóvel’ está claramente identificado. No entanto, obtém-se informação pertinente que nos

<sup>4</sup> in Dicionário da Língua Portuguesa com Acordo Ortográfico [em linha]. Porto: Porto Editora, 2003-2016. [consult. 2016-03-08 00:06:33]

<sup>5</sup> in Dicionário Priberam da Língua Portuguesa [em linha], 2008-2013, <https://www.priberam.pt/DLPO/seguranca> [consultado em 08-03-2016]

permite identificar os termos que podem constituir possíveis campos lexicais que poderão vir a ser úteis para constituir a terminologia associada ao conceito <segurança automóvel>. No caso do termo ‘automóvel’, podemos reter outros termos associados, tal como ‘veículo’ que, podemos adiantar, é o hiperónimo de ‘automóvel’. O mesmo raciocínio pode ser desenvolvido para o termo ‘segurança’.

## 1.2. Consulta de sítios

No sentido de se avançar com a descrição do domínio, recorremos a um motor de busca para pesquisar o termo ‘segurança automóvel’. Os resultados que obtivemos são os seguintes:

**Figura 3– Resultado da pesquisa com motor de busca**



Como se pode verificar na Figura 3, os resultados que se obtém são de vária ordem: ligações para documentos avulsos em formato PDF, ligações para fóruns sobre assuntos variados, sítios de fabricantes de automóveis e ligações a entidades governamentais, nacionais e internacionais.

Da lista dos resultados obtidos, selecionámos 3 hiperligações, que correspondem aos seguintes resultados dispostos na Figura3:

Hiperligação 1:

<http://www.ansr.pt/SegurancaRodoviaria/ArtigosTecnicos/Documents/Seguran%C3%A7a%20Passiva%20Autom%C3%B3vel.pdf>, corresponde ao resultado “[PDF]Segurança Passiva Automóvel – Ansr”;

Hiperligação 2:

[http://www.imtt.pt/sites/IMTT/Portugues/EnsinoConducao/ManuaisEnsinoConducao/Documents/Fichas/FT\\_SistemasSegurancaPassiva.pdf](http://www.imtt.pt/sites/IMTT/Portugues/EnsinoConducao/ManuaisEnsinoConducao/Documents/Fichas/FT_SistemasSegurancaPassiva.pdf), corresponde ao resultado “[PDF]sistemas de segurança passiva – Imtt”;

Hiperligação 3:

[http://ec.europa.eu/transport/road\\_safety/topics/vehicles/esafety/index\\_pt.htm](http://ec.europa.eu/transport/road_safety/topics/vehicles/esafety/index_pt.htm), corresponde ao resultado “Sistemas de segurança inteligentes para os veículos...”

Estas três hiperligações remetem para tipos de suportes diferentes. A duas primeiras remetem para documentos PDF. A última para um sítio.

Passamos, assim, à descrição destas 3 hiperligações:

### 1.2.1 Hiperligação 1

A opção pela hiperligação 1 justifica-se pelo facto de a primeira informação que aparece depois de a ativar, conter as duas formas que compõem o termo que designa o domínio, a saber ‘segurança’ e ‘automóvel’. Esta informação encontra-se num documento em formato “PDF” pertencente à Autoridade Nacional de Segurança Rodoviária (ANSR). O documento intitula-se “Segurança Passiva Automóvel” e o autor é Diogo Júdice – Técnico Superior. Para além destes dados, o documento tem o logotipo da ANSR nas suas páginas, mas nada mais que nos indique a sua real proveniência e quais as competências do autor para o escrever.

Estes resultados levaram-nos a consultar a página principal da ANSR<sup>6</sup> na tentativa de clarificar quais as reais competências desta entidade. Aqui verificámos que os conteúdos principais se encontram organizados num índice. O primeiro dos conteúdos contém informação acerca da ANSR (missão, colaboradores...), o que nos permite ficar mais

---

<sup>6</sup> <http://www.ansr.pt/Pages/default.aspx>



esclarecidos quanto à sua missão, levando-nos de seguida a uma outra hiperligação<sup>7</sup> que nos encaminha para uma página intitulada “A ANSR”, onde é descrita a sua missão:

“A Autoridade Nacional de Segurança Rodoviária (ANSR) é um serviço central da administração direta do Estado dotado de autonomia administrativa.

A ANSR tem por missão o planeamento e coordenação a nível nacional de apoio à política do Governo em matéria de segurança rodoviária, bem como a aplicação do direito contraordenacional rodoviário”

Ainda nesta página, destacamos algumas das principais atribuições da ANSR:

“Contribuir para a definição das políticas no domínio do trânsito e da segurança rodoviária;

Elaborar e monitorizar o Plano Nacional de Segurança Rodoviária, bem como os documentos estruturantes relacionados com a segurança rodoviária, e bem assim promover o seu estudo, nomeadamente das causas e fatores intervenientes nos acidentes de trânsito;”

É de notar que o sítio da ANSR disponibiliza aos utentes das vias públicas, peões e automobilistas abundante informação relacionada com automóveis, campanhas de sensibilização e legislação.

Após esta incursão, voltámos à página principal<sup>8</sup>, para atendermos ao segundo conteúdo apresentado no índice, intitulado, “SEGURANÇA RODOVIÁRIA”. Associado à temática da “segurança rodoviária”, encontramos os seguintes itens:

- História
- Código da Estrada
- Planos de segurança Rodoviária
- Campanhas
- Publicações
- Conselhos
- Informação Técnica
- Artigos Técnicos
- Internacional
- Regulamento de Sinalização de Trânsito

A seleção de qualquer um dos itens atrás mencionados, redireciona-nos para uma página, onde são disponibilizados documentos técnicos e textos legais. A título

---

<sup>7</sup> <http://www.ansr.pt/AANSR/Pages/default.aspx>

<sup>8</sup> <http://www.ansr.pt/Pages/default.aspx>

exemplificativo, ao selecionar o item “Código da Estrada”, somos direcionados para uma página<sup>9</sup>, onde podemos consultar uma versão atualizada do Código da Estrada, que pode ser descarregado. No interior da ligação identificada como “Publicações” existem diversos documentos. Um deles intitula-se “Melhores práticas de segurança rodoviária” e é um documento que contém sugestões em matéria de segurança rodoviária, acessível também para consulta, com possibilidade de ser descarregado<sup>10</sup>.

### 1.2.2. Hiperligação 2

A hiperligação 2 que também remete para um PDF, intitula-se “sistemas de segurança passiva”. Optámos por consultá-lo, por analogia à informação contida no PDF anterior que, recordamos, diz respeito à “Segurança Passiva Automóvel”, o que nos leva a deduzir que este PDF intitulado “sistema de segurança passiva” diz respeito ao domínio em análise.

O documento para o qual esta ligação remete é um excerto que faz parte de um manual de ensino da condução. Não é possível determinar com segurança de que documento este excerto faz parte; não tem indicação de nenhum autor em particular, apenas traz a o logotipo do IMT nas páginas do documento.

Por esta via, fomos levados a consultar a página principal do IMT<sup>11</sup> na tentativa de clarificar quais as reais competências desta entidade. Assim, verificámos que os conteúdos principais estão organizados num índice localizado na parte lateral esquerda da página. Aí, ao selecionarmos o segundo conteúdo, designado “O IMT”<sup>12</sup>, podemos encontrar a seguinte informação acerca da entidade:

“O Instituto da Mobilidade e dos Transportes (IMT, I.P.) é um instituto público integrado na administração indireta do Estado, dotado de autonomia administrativa e financeira e património próprio.”

“O IMT, I.P. é um organismo central com jurisdição sobre todo o território nacional, tem sede em Lisboa e dispõe, como serviços desconcentrados, das Direções Regionais de Mobilidade e Transportes do Norte, do Centro, de Lisboa e Vale do Tejo, do Alentejo e do Algarve.”

---

<sup>9</sup> <http://www.ansr.pt/SegurancaRodoviaria/CodigoDaEstrada/Pages/default.aspx>

<sup>10</sup> <http://www.ansr.pt/SegurancaRodoviaria/Publicacoes/Documents/Melhores%20práticas%20de%20Segurança%20Rodoviária.pdf>

<sup>11</sup> <http://www.imtt.pt/sites/IMTT/Portugues/Paginas/IMTHome.aspx>

<sup>12</sup> <http://www.imtt.pt/sites/IMTT/Portugues/IMTT/Paginas/OIMT.aspx>

Ao clicarmos no item “O IMT”, fomos direcionados para a “Missão e Atribuições”<sup>13</sup>, onde encontramos descritas as suas principais missões:

“O exercício das funções de regulamentação técnica, de licenciamento, coordenação, fiscalização e planeamento no setor dos transportes terrestres, fluviais e respetivas infraestruturas e na vertente económica do setor dos portos comerciais e transportes marítimos”

Ainda dentro do mesmo item, destacamos de entre as suas atribuições, três que nos parecem relevantes para a análise aqui desenvolvida:

- (1) “Promover os aperfeiçoamentos técnicos em veículos rodoviários (...)”;
- (2) “ (...) com o objetivo de melhorar a segurança e a eficiência da exploração dos transportes rodoviários e ferroviários, a interoperabilidade e a redução de impactos ambientais negativos”;
- (3) “Aprovar, homologar e certificar veículos e equipamentos afetos aos sistemas de transporte terrestre, garantindo os padrões técnicos e de segurança exigidos (...)”

A navegação que efetuámos pelos vários menus da página principal permitiu constatar quais as principais áreas de atividade do IMT e como elas se encontram estruturadas.

### 1.2.3 Hiperligação 3

A opção pela hiperligação 3 está relacionada com o facto de o título mencionar “sistemas de segurança inteligentes para os veículos”. Constatamos depois tratar-se especificamente de veículos do tipo automóvel, o que parece relacionar-se com o domínio.

A hiperligação conduz diretamente a um sítio pertencente à Comissão Europeia (CE) onde nos é dito o que se entende por “Sistemas de segurança inteligentes para os veículos automóveis”.

“Os sistemas de segurança inteligentes para os veículos automóveis são frequentemente designados sistemas “eSafety”. Trata-se de sistemas electrónicos concebidos para ajudar os condutores de veículos automóveis a evitar situações de perigo, por exemplo, activando os travões se o automóvel se aproximar demasiado de um objecto ou emitindo sinais de alarme se sair da faixa de rodagem.”

A nossa atenção vira-se para a estrutura do índice desta página que está localizado na parte lateral esquerda da mesma. Podemos perceber que o texto atrás mencionado se encontra

---

13

<http://www.imtt.pt/sites/IMTT/Portugues/IMTT/MissaoeAtribuicoes/Paginas/MissaoeAtribuicoes.aspx>

na ponta final de uma sucessão de temas com início na página principal da CE, seguida do tema dos transportes; da segurança rodoviária; dos veículos automóveis e terminando na página que alberga o texto atrás citado.

A análise desta página leva a que visitemos a página principal da CE <sup>14</sup>, com o intuito de obter mais informações acerca desta entidade e determinar quais as suas principais competências e funções.

A visita inicia-se com uma hiperligação intitulada “saiba mais sobre a comissão”, que nos redireciona para a página <sup>15</sup> onde se pode ler:

“A Comissão Europeia é o órgão executivo da União Europeia, representando os seus interesses no conjunto (e não os interesses específicos de cada país)”

Aí também se encontram as suas principais funções;

(1) “propor legislação que é em seguida adotada pelos legisladores, ou seja, o Parlamento Europeu e o Conselho de Ministros”;

(2) “assegurar o cumprimento do direito europeu (se necessário, com a ajuda do Tribunal de Justiça da União Europeia)”

Parece-nos evidente, em função das competências atrás referidas e, pelo que pudemos encontrar na página inicialmente consultada, que a CE tem autoridade e interesse em atuar em matéria de segurança automóvel.

Apesar de na pesquisa efetuada termos obtido mais resultados do que aqueles que apresentamos aqui para o termo ‘segurança automóvel’, consideramos ter já neste ponto dados satisfatórios para uma descrição do domínio.

### **1.3. Áreas de atividade das entidades: CE, IMT, ANSR**

A escolha pelas hiperligações 1, 2 e 3 justifica-se pela presença dos termos simples constituintes do termo complexo ‘segurança automóvel’ e de termos com ele conexos. Para além deste critério, acrescenta o facto de as hiperligações apontarem para entidades com competências no sentido de criar e aplicar legislação no domínio da Segurança Automóvel em termos nacionais, no caso do IMT e da ANSR, no caso da CE, em termos internacionais.

---

<sup>14</sup> [http://ec.europa.eu/index\\_pt.htm](http://ec.europa.eu/index_pt.htm),

<sup>15</sup> [http://ec.europa.eu/about/index\\_pt.htm](http://ec.europa.eu/about/index_pt.htm)

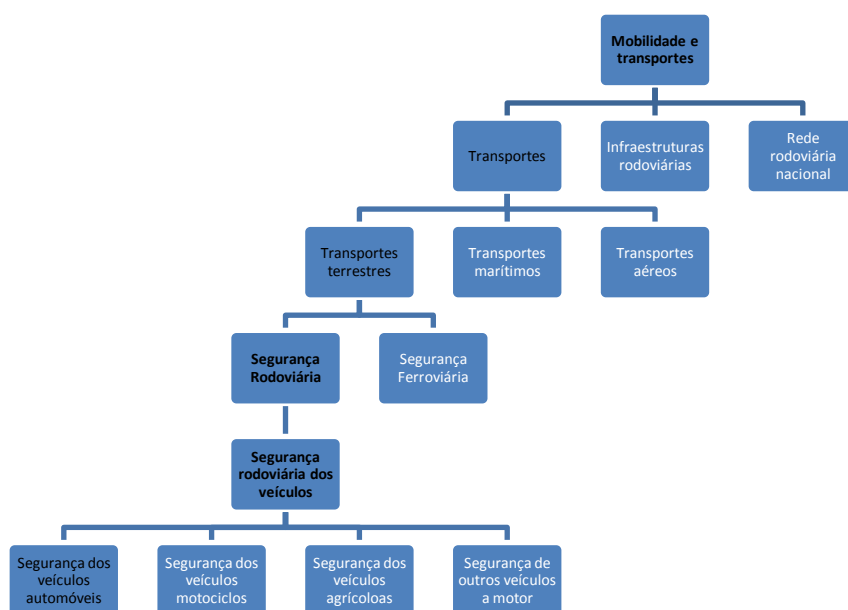
A observação das páginas das entidades atrás referidas possibilitou identificar quais as suas principais áreas de atividade no que diz respeito ao domínio da Segurança Automóvel, pelo que decidimos agrupá-las da seguinte forma:

**Tabela 2**

Comissão Europeia	IMT	ANSR
<b>transportes</b> <b>segurança rodoviária</b> <b>veículos automóveis</b>	mobilidade e transportes terrestres infraestruturas rodoviárias rede rodoviária nacional	código da estrada planos de segurança rodoviária regulamento de sinalização de trânsito

Com base na observação que fizemos das páginas principais da CE, do IMT e da ANSR, propomos a seguinte árvore de domínio, assinalando a negrito, os itens que em nosso entender constituem o domínio da Segurança Automóvel e sobre os quais irá incidir o nosso trabalho.

**Figura 4 - Estrutura do domínio**



## 1.4. Descrição dos conceitos de base do domínio

Em resultado do domínio atrás representado, julgamos ser importante dissertar sobre três possíveis conceitos que destacámos a negrito na figura anterior: o de <segurança rodoviária>, o de <segurança rodoviária dos veículos> e o de <segurança dos veículos automóveis>.

Começamos por analisar o conceito de <segurança rodoviária>, recorrendo a um documento disponível no sítio da CE<sup>16</sup> intitulado “Rumo a um espaço europeu de segurança rodoviária”. Este documento contém um conjunto de propostas relativas à segurança rodoviária cujo propósito é o de contribuir para a redução do número de vítimas de acidentes rodoviários na UE.

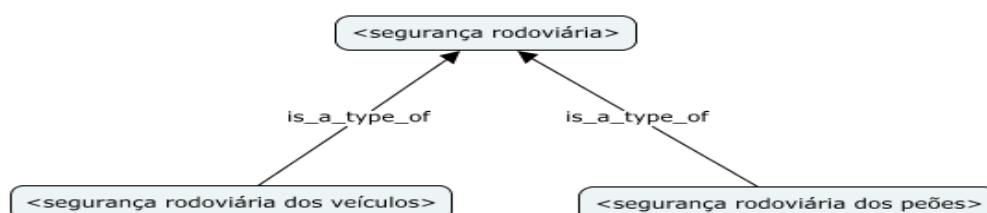
A consulta deste documento permite verificar que falar em segurança rodoviária pressupõe um conjunto de políticas de âmbito internacional cujo objetivo é:

“aumentar a segurança na estrada, proporcionando uma mobilidade segura e ecológica aos cidadãos de toda a Europa. Deverá promover a equidade dos utentes da estrada, centrando esforços na melhoria da segurança dos utentes mais vulneráveis”

A segurança rodoviária caracteriza-se por um conjunto de práticas, levadas a cabo por governos, autoridades e grupos de interesse, no intuito de melhorar a segurança das estradas, veículos e pessoas que nelas circulam.

Apenas com base na interpretação do documento disponível no sítio da CE (ver Nota de rodapé 16) não é possível determinar conclusivamente o que se entende por <segurança rodoviária>. No entanto, no contexto em que ocorre o conceito expresso, podemos encontrar fornecem informação acerca de outros termos associados ao termo que designa o conceito de <segurança automóvel>. A partir desta informação, podemos afirmar que existe uma relação de subordinação do conceito <segurança rodoviária dos veículos> em relação ao conceito subordinante <segurança rodoviária>, que corresponde à seguinte representação:

### Mapa conceptual 1



Os conceitos designados pelos termos ‘segurança rodoviária’ e ‘segurança rodoviária dos veículos’ estabelecem entre si uma relação genérico – específico.

No caso de <segurança dos veículos automóveis>, consideramos ser importante analisar isoladamente os termos ‘veículos’ e ‘automóveis’, que compõem o termo complexo ‘veículos automóveis’.

<sup>16</sup> [http://ec.europa.eu/transport/road\\_safety/pdf/road\\_safety\\_citizen/road\\_safety\\_citizen\\_100924\\_pt.pdf](http://ec.europa.eu/transport/road_safety/pdf/road_safety_citizen/road_safety_citizen_100924_pt.pdf)

No ponto 2 do documento atrás citado, podemos ler que a CE inclui no seu programa de ação, medidas de segurança especificamente destinadas aos veículos automóveis. Em função destas medidas, a nossa interpretação é de que a segurança dos veículos automóveis é uma área que está sob a alçada da Comissão europeia. Contudo, também existem entidades que aplicam medidas semelhantes a nível nacional e, no sentido de esclarecer o conceito de <veículos automóveis>, optamos por analisar a definição que está contida no texto do Código da Estrada- lei n.º 72/2013<sup>17</sup>.

Assim, podemos verificar que, de acordo com o artigo 105º do Código da Estrada- lei n.º 72/2013, <automóvel> é definido como

“o veículo com motor de propulsão, dotado de pelo menos quatro rodas, com tara superior a 550 kg, cuja velocidade máxima é, por construção, superior a 25 km/h, e que se destina, pela sua função, a transitar na via pública, sem sujeição a carris”

Esta definição pode ser consultada no TÍTULO IV, CAPÍTULO I do Código da Estrada- lei n.º 72/2013, intitulado de “Classificação dos veículos” e, em função da mesma verifica-se que um automóvel é um tipo específico de veículo.

Os artigos seguintes referem-se à forma como os veículos automóveis se encontram organizados em função das diferentes classes e tipos que existem. Ao observarmos o artigo 106º, por exemplo, verificamos a distinção que é estabelecida nas suas alíneas, entre os vários tipos de automóveis. Assim,

os automóveis classificam-se em:

- a) Ligeiros - veículos com peso bruto igual ou inferior a 3500 kg e com lotação não superior a nove lugares, incluindo o do condutor;
- b) Pesados - veículos com peso bruto superior a 3500 kg ou com lotação superior a nove lugares, incluindo o do condutor.”

Este artigo acrescenta informação que determina a existência de diferentes tipos de veículos automóveis, que se distinguem em função de características específicas, como por exemplo o peso.

Para além da classificação anterior do conceito de <veículos automóveis>, disposta no Código da Estrada, importa analisar uma outra designação para o conceito atrás mencionado, atribuída ao IMT.

---

17

[http://www.ansr.pt/SegurancaRodoviaria/CodigoDaEstrada/Documents/Codigo\\_Estrada\\_2014\\_versao\\_WEB.pdf](http://www.ansr.pt/SegurancaRodoviaria/CodigoDaEstrada/Documents/Codigo_Estrada_2014_versao_WEB.pdf)

A designação utilizada pelo IMT serve para ser utilizada em ações de certificação dos veículos. Esta certificação atesta a conformidade de um modelo automóvel com os requisitos técnicos estabelecidos na legislação e designa-se por ‘homologação’. A classificação para efeitos de homologação é de âmbito europeu e encontra-se disposta da seguinte forma:

**Figura 5 – Classificação europeia de veículos**

Categoria	Definição
<b>M</b> Veículos a motor concebidos e construídos para o transporte de passageiros com, pelo menos, quatro rodas	<b>M1</b> - Veículos concebidos e construídos para o transporte de passageiros com oito lugares sentados no máximo, além do lugar do condutor
	<b>M2</b> - Veículos concebidos e construídos para o transporte de passageiros, com mais de oito lugares sentados além do lugar do condutor e uma massa máxima não superior a 5 t
	<b>M3</b> - Veículos concebidos e construídos para o transporte de passageiros, com mais de oito lugares sentados além do condutor e uma massa máxima superior a 5 t
<b>N</b> Veículos a motor concebidos e construídos para o transporte de mercadorias com, pelo menos, quatro rodas	<b>N1</b> - Veículos concebidos e construídos para o transporte de mercadorias com massa máxima não superior a 3,5 t
	<b>N2</b> - Veículos concebidos e construídos para o transporte de mercadorias com massa máxima superior a 3,5 t mas não superior a 12 t
	<b>N3</b> - Veículos concebidos e construídos para o transporte de mercadorias com massa máxima superior a 12 t
<b>O</b> Reboques (incluindo os semi-reboques)	<b>O1</b> - Reboques com massa máxima não superior a 0,75 t
	<b>O2</b> - Reboques com massa máxima superior a 0,75 t mas não superior a 3,5 t
	<b>O3</b> - Reboques com massa máxima superior a 3,5 t mas não superior a 10 t
	<b>O4</b> - Reboques com massa máxima superior a 10 t

Por via da informação que nos é dada na Figura 5, não nos é possível determinar qual a designação do conceito de <automóvel>. A leitura que fazemos da Figura leva-nos a concluir que o termo que designa o conceito de <automóvel> está condicionado pelo uso que diferentes entidades ligadas ao domínio da segurança automóvel lhe dão.

O termo ‘automóvel’ parece ser utilizado como designação genérica para o conceito de <automóvel>, mas a interpretação que fazemos da informação não é suficiente para eliminar as dúvidas. Seria necessária a assistência de um especialista.

No caso de ‘veículo’, consideramos que a sua combinação com ‘automóvel’ restringe a designação do conceito de <automóvel>, limitando-o a um tipo específico de veículo, correspondendo esta combinação também a uma designação genérica, pois como verificámos anteriormente, existem diversos tipos de veículos automóveis.



### 1.4.1. Mapas conceptuais

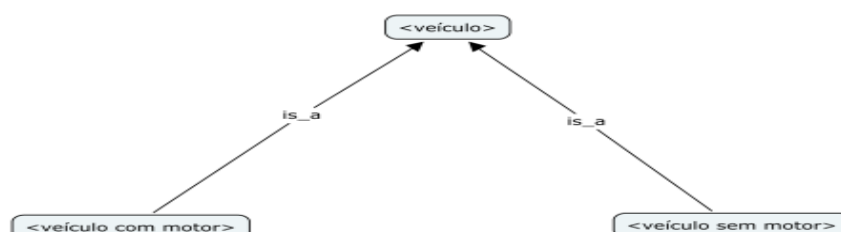
Apresentamos de seguida uma proposta de organização dos conceitos atrás discriminados, dispondo-os de acordo com as suas características, num mapa conceptual:

A nossa análise ao TÍTULO IV do Código da Estrada e à página do IMT <sup>18</sup>, de onde se retirou a imagem que conta na Figura anterior, leva-nos a concluir que existem mais do que uma designação para o conceito de <veículo automóvel> e que existe a possibilidade de este ser representado por diferentes termos.

Com base nesta análise colocamos a hipótese de o mesmo se poder verificar com o conceito de <segurança automóvel>. É nossa intenção verificar tal facto durante a tarefa de análise do *corpus*.

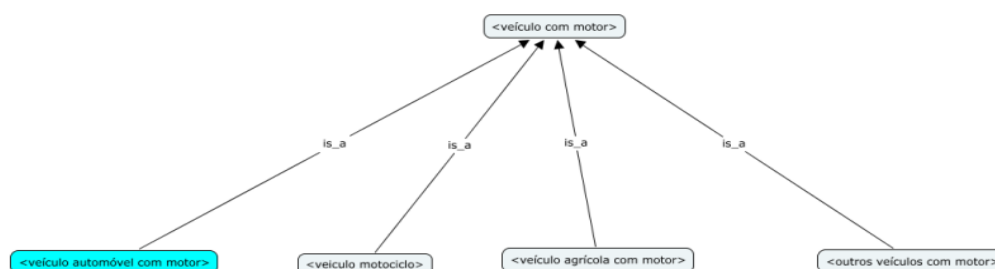
Passamos assim a apresentar a proposta de organização do domínio. O Mapa conceptual 2 corresponde aos tipos de <veículos> que existem:

#### Mapa conceptual 2



O <veículo com motor> subordina os conceitos seguintes:

#### Mapa conceptual 3



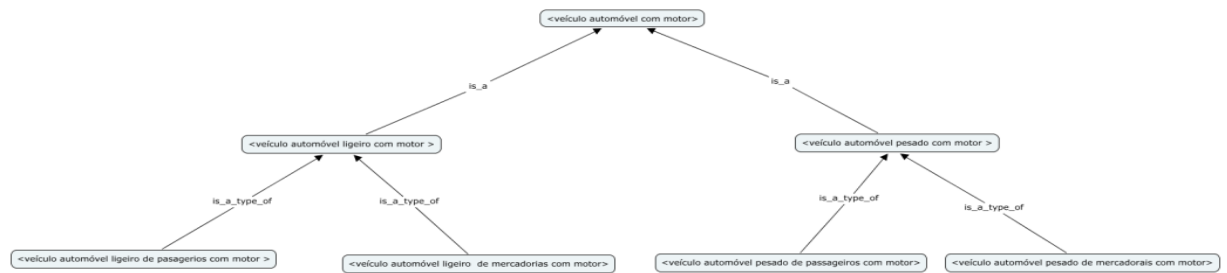
A partir deste mapa, iremos focalizar-nos no conceito de <veículo automóvel com motor> que representamos como se segue:

---

18

<http://www.imtt.pt/sites/IMTT/Portugues/Veiculos/Aprovacoes/HomologacoesVeiculos/Paginas/HomologacaoVeiculos.aspx>

## Mapa conceptual 4



Com base nestes mapas, verificámos que <segurança dos veículos automóveis> tanto dizem respeito a <veículos com motor>, como a <veículos sem motor>. O conceito de <automóvel> é descrito da seguinte forma no Código da Estrada, Artigo 105:

“Automóvel é o veículo com motor de propulsão, dotado de pelo menos quatro rodas, com tara superior a 550 kg, cuja velocidade máxima é, por construção, superior a 25 km/h, e que se destina, pela sua função, a transitar na via pública, sem sujeição a carris.”

Como podemos verificar, neste conceito mais genérico, o conceito <automóvel> é descrito como sendo um “veículo com motor de propulsão”, não abrangendo <veículo sem motor>, a saber, por exemplo, reboques e velocípedes. Pensamos, no entanto, que <segurança dos veículos automóveis> também abarca estes últimos. Estamos assim perante uma ambiguidade que seria importante resolver. No entanto, esta questão não será tratada no âmbito deste trabalho por necessitarmos da validação de peritos na matéria.

## 2. Linguística de *corpus*

A linguística de *corpus* tem como objetivo, estudar o funcionamento da língua, recorrendo para tal, à análise de grandes quantidades de textos em formato eletrónico, organizados em função de critérios específicos, os *corpora*.

Os linguistas que tentam explicar o funcionamento de uma língua, começam por definir as suas hipóteses com base na sua intuição acerca do funcionamento da mesma. O recurso aos *corpora* permite que os linguistas confirmem as suas hipóteses, originadas pela introspeção, com dados de língua real.

Como tal, a linguística de *corpus* beneficiou com o aparecimento da informática e o seu posterior desenvolvimento. A facilidade de análise de grandes quantidades de textos, assim como a sua organização e armazenamento, decorre do aparecimento de ferramentas de tratamento semiautomático da língua, desenhadas com o propósito de auxiliar na análise aos *corpora*. A utilização destas ferramentas permite diminuir o tempo gasto a analisar um *corpus* de grandes dimensões em busca da informação de que se necessita, possibilitando fazer análises mais rápidas e eficazes.

No campo da linguística de *corpus*, podemos constatar que existem atividades que se destacam, como a compilação e desenho de *corpora*, consistindo este trabalho na seleção, recolha, preparação, organização e armazenamento de textos; o desenvolvimento de ferramentas de análise de *corpora*; o uso de *corpora* eletrónicos com o intuito de descrever o léxico e a gramática das línguas; o uso e aplicação das metodologias de análise de *corpora* a outras atividades relacionadas com a língua, como o ensino e aprendizagem de línguas, processamento da língua natural, tradução automática, etc.

A linguística de *corpus* assenta nos textos que são alvo de análise e estudo, fornecendo auxílio à descrição de uma língua quanto à quantificação e distribuição dos itens linguísticos que a compõem. A principal razão para a compilação de *corpora* linguísticos é fornecer uma base para uma descrição da forma como as línguas se estruturam e são usadas, com aplicações em matéria de ensino de línguas, na lexicografia, no auxílio da tradução e na área da terminologia.

## 2.1. *Corpora* e ensino da língua

Na altura da conceção de metodologias de análise de *corpora*, aplicáveis ao ensino de língua de especialidade, consideramos importante destacar dois métodos de utilização de *corpora*, que apesar de utilizarem o mesmo recurso, o *corpus*, apresentam algumas diferenças quanto aos seus objetivos e metodologias.

Gavioli (2005, p.23) refere a existência de duas tendências na realização de estudos sobre a utilização de *corpora* no ensino de línguas. A primeira tem como foco analisar o uso de *corpora* por parte dos professores de línguas, a segunda, tem como objetivo determinar os benefícios obtidos por aprendentes de línguas ao usar *corpora*.

Para esta autora “*A corpus of specialized texts can be used to rapidly generate a series of items which may characterize the specialized texts in question.*” (2005, p.23), ou seja, para o terminólogo, os *corpora* de especialidade funcionam como um repositório de onde se pode extrair os termos que compõem o domínio de especialidade em observação.

Um *corpus* é um recurso rico em dados reais, cuja autenticidade possibilita aos alunos interagir com um *corpus*, aprendam por si próprios um uso possível da língua.

Nos casos específicos em que os *corpora* são constituídos por produções de falantes nativos da língua em aprender, permite aos aprendentes constatar de que forma uma palavra ou construção gramatical ocorre tipicamente numa língua.

Cada vez mais a investigação com base na análise de *corpora* é importante na medida em que pode contribuir para o desenvolvimento de estratégias de ensino.

Uma destas estratégias é ensinar os alunos a utilizar ferramentas de tratamento semiautomático da língua, como os concordanceiros sendo que estes permitem analisar um *corpus* de forma aprofundada em função de objetivos específicos de aprendizagem da língua. A extração de todas as formas que ocorrem num *corpus*, para depois as organizar por ordem de frequência, pode apontar para contextos também eles frequentes, contribuindo para a aprendizagem.

## 2.2. *Corpora* e terminologia

A terminologia é uma disciplina que beneficiou das vantagens providenciadas pela automatização dos processos de compilação e análise de *corpora*, sendo que, Sager (1990, p.

129) reconhece “ (...) *that the only practical means of processing lexical data is by computer*”.

De acordo com Sager (1990, p.130), um *corpus* é “(...) *a representative body of texts of a subject field which in this way is confined in a very concrete way*”. Os *corpora*, para fins terminológicos, destacam-se pela natureza dos textos que os compõem. Estes textos são redigidos por especialistas e representam o conhecimento relativo a um domínio especializado.

Constituir um *corpus* de especialidade implica a definição de outros processos que têm a ver diretamente com a delimitação do próprio domínio de especialidade a que os textos devem pertencer, como defende l’Homme (2004, p.119):

“La collecte d'une documentation représentative du domaine dont on souhaite décrire la terminologie et son exploitation constituent les premières étapes d'une recherche en bonne et due forme.”

Uma outra característica é que os *corpora* são condicionados em função do período temporal a que os textos pertencem, sendo necessário uma atualização constante, pois segundo Sager (1990, p.131), “*A certain item of information extracted from a particular corpus at a particular time is valid as long as the textual basis maintains its actuality.*”

Os *corpora* especializados não se limitam a acentuar aspetos particulares de uma língua. Eles também permitem identificar comportamentos lexicais que apontam para conhecimento extralinguístico específico a um domínio. Cabe ao terminólogo, após análise dos *corpora*, extrair as características principais que aí ocorrem, permitindo a compilação de glossários e a construção de bases de dados terminológicas, sejam elas textuais ou relacionais.

### 3. Constituição do *corpus*

Um ponto importante a discutir antes de se iniciar a tarefa de constituição de um *corpus*, é saber que textos se devem incluir por forma a cumprir os objetivos estipulados pelos investigadores.

Para levar a cabo a nossa pesquisa, optámos por constituir um *corpus* monolíngue, tendo sido a língua escolhida a portuguesa, pois estabelecemos que os futuros utilizadores do *corpus* tenham um nível médio a avançado de proficiência em português L2.

Os textos escritos existem em grande quantidade em formato digital, o que torna mais fácil a tarefa de reunir recursos especializados, possibilitando constituir um *corpus* de maior dimensão.

No que diz respeito ao tamanho do *corpus*, autores como Sinclair (1991) são da opinião que este deve ser o maior possível, outros como Bowker e Pearson (2002) não concordam com esta adjetivação, consideram-na vaga, referindo que não é possível enumerar regras que possam ser aplicadas e que determinem o tamanho ideal de um *corpus*.

No caso do nosso *corpus* podemos apontar, como fatores que condicionam o seu tamanho, a existência de textos disponíveis e o facto de estes serem recentes e redigidos por especialistas. É preferível que um *corpus* seja pequeno e bem desenhado, por oposição a um *corpus* de grande dimensão, mas que não se adequa aos objetivos da investigação.

Durante o processo de descrição do domínio, patente no primeiro capítulo deste trabalho, visitámos vários sítios pertencentes a entidades com competências relativas ao domínio da Segurança Automóvel, a saber, os sítios da ANSR, do IMT e da CE, tendo escolhido esta última como fonte dos documentos para o nosso *corpus*, pelos motivos que apresentamos de seguida:

A consulta ao sítio da CE revelou uma série de ligações relacionadas com a segurança automóvel. A exploração de diversas dessas ligações encaminhou-nos para documentos em PDF<sup>19</sup>, cuja origem viemos a determinar como sendo a base de dados EUR-lex.

A base de dados EUR-lex é um recurso oficial integrado no campo administrativo da CE que é atualizada diariamente, fornecendo desse modo documentação recente, disponível de forma livre e gratuita, proveniente de uma instituição de prestígio comprovado e redigida por especialistas.

---

<sup>19</sup> <http://eur-lex.europa.eu/legal-content/PT/ALL/?uri=CELEX:32005L0039>

A escolha da CE como fonte de documentos para o nosso *corpus*, resulta do facto de a regulamentação da segurança automóvel ser da responsabilidade da comissão e a legislação ser aplicada de igual forma a todos os estados membros.

### 3.1. Seleção dos textos

Os documentos estão, na sua maioria, relacionados com o Direito na União Europeia (EU), na forma de diretivas, regulamentos, decisões e tratados. Os documentos estão disponíveis em diferentes formatos de texto e em várias línguas. Constatámos, contudo, no decurso da nossa pesquisa, algumas exceções referentes a textos que apenas estão acessíveis em Inglês.

Para proceder à escolha do tipo de documentos a integrar o *corpus*, visitámos o sítio oficial da União Europeia<sup>20</sup>. Uma vez lá, seleccionámos o tema relativo ao direito da EU<sup>21</sup>, onde fomos descobrir a informação referente ao tipo de documentos disponível.

Na página anteriormente referida verificamos que o direito na EU está dividido em direito primário; “Os tratados (direito primário) constituem as regras fundamentais que estão na base de todas as medidas tomadas pela UE.” e, direito derivado; “constituído por regulamentos, diretivas e decisões, decorre dos princípios e objetivos estabelecidos nos tratados.”.

O nosso interesse foca-se nos documentos pertencentes ao direito derivado, a saber, diretivas, decisões, regulamentos recomendações e pareceres.

Deste lote de possíveis escolhas, optámos pelos regulamentos, pelo facto de estes serem documentos vinculativos e aplicáveis em todos os países da EU, oferecendo garantia quanto ao cumprimento dos seus pressupostos em território português.

Para a tarefa de recolha de textos começámos por estabelecer, com base na informação encontrada na página inicial da EUR-lex<sup>22</sup>, os critérios que serviriam de guia à pesquisa.

Em primeiro lugar, estabelecemos o Português como língua de pesquisa, seguida dos seguintes critérios:

---

<sup>20</sup> [http://europa.eu/index\\_pt.htm](http://europa.eu/index_pt.htm)

<sup>21</sup> [http://europa.eu/eu-law/index\\_pt.htm](http://europa.eu/eu-law/index_pt.htm)

<sup>22</sup> <http://eur-lex.europa.eu/homepage.html>

- Domínio: Legislação da UE e documentos conexos,
- Subdomínio: Legislação,
- Autor: Comissão Europeia,
- Forma: regulamento,
- Resultados que contêm: segurança automóvel No título e texto,
- OR (ou): rodoviária No título e texto,
- OR (ou): veículos automóveis No título e texto

No campo referente à data, estabelecemos um período de dez anos quanto à data de publicação dos documentos, desde janeiro de 2005 até dezembro de 2015 inclusive.

O intervalo de dez anos permite, ao consultar os documentos do período inicial, estabelecer uma comparação entre estes e os mais recentes, tornando possível observar a manifestação de novos conceitos e as suas definições, assim como acompanhar a sua evolução ao longo do período escolhido. Este período justifica-se por se considerar que este representa a atualidade do domínio da Segurança Automóvel, tanto a nível linguístico como conceptual.

Um outro motivo é este intervalo permitir o acesso a versões anteriores e aumentar o número possível de documentos a consultar, contribuindo para a nossa familiarização com o domínio.

O último critério foi selecionar documentação produzida pela Comissão Económica das Nações Unidas para a Europa (UNECE).

Foi possível obter, a partir destes critérios de pesquisa, 80 resultados, organizados por data e em sentido descendente.

Na Figura 6 podemos ver um exemplo do ambiente de pesquisa que surge ao escolhermos um texto. Neste ambiente de trabalho temos à disposição uma série de opções, permitindo obter informação adicional sobre o texto selecionado.

Por exemplo, ao selecionar o campo intitulado “Documentos conexos”, somos redirecionados para uma área onde temos à disposição uma série de *links*. Um desses *links*<sup>23</sup>, intitulado “Visualizar o Jornal Oficial onde se encontra o documento”, redireciona-nos para o Jornal Oficial da União Europeia, onde se encontra publicado o texto que estamos a observar:

---

<sup>23</sup> <http://eur-lex.europa.eu/legal-content/PT/TXT/?uri=OJ:L:2007:121:TOC>



**Figura 6 – Menu de trabalho da base de dados EUR-lex**

Documento 32004L0104 > [Guardar em «Os meus resultados»](#) [Adicionar este elemento aos favoritos](#) [Descarregar a ficha](#)

Acerca deste documento	Texto	Documentos conexos	Tudo	MNE	Fechar tudo   Mostrar tudo
<b>Título e referência</b> Directiva 2004/104/CE da Comissão, de 14 de Outubro de 2004, que adapta ao progresso técnico a Directiva 72/245/CEE do Conselho relativa às interferências radioeléctricas (compatibilidade electromagnética) dos veículos e que altera a Directiva 70/156/CEE relativa à aproximação das legislações dos Estados-Membros respeitantes à homologação dos veículos a motor e seus reboques <i>JO L 337 de 13.11.2004, p. 13—58 (ES, CS, DA, DE, ET, EL, EN, FR, IT, LV, LT, HU, NL, PL, PT, SK, SL, FI, SV)</i> <i>JO L 183M de 5.7.2006, p. 283—328 (MT)</i> <i>edição especial em língua búlgara: Capítulo 13 Fascículo 045 p. 201 – 246</i> <i>edição especial em língua romena: Capítulo 13 Fascículo 045 p. 201 – 246</i> <i>edição especial em língua croata: Capítulo 13 Fascículo 024 p. 80 – 125</i>					
<b>Língua que faz fé</b> • Língua que faz fé: espanhol, checo, dinamarquês, alemão, estónio, grego, inglês, francês, italiano, letão, lituano, húngaro, maltês, neerlandês, polaco, português, eslovaco, esloveno, finlandês, sueco, islandês, norueguês					
<b>Datas</b> • Data do documento: 14/10/2004 • Data de efeito: 03/12/2004; entrada em vigor data de publicação + 20 ver art. 5 • Data de efeito: 03/12/2004; entrada em vigor data de publicação + 20 ver art. 5 • Data de transposição: 31/12/2005; o mais tardar até ver art. 4 • Data de transposição: 31/12/2005; o mais tardar até ver art. 4 • Data do termo de validade: 31/12/9999					
<b>Classificações</b> • Descritor EUROVOC: aproximação das legislações matrícula do veículo veículo motorizado perturbação eletromagnética mudança tecnológica marca de conformidade CE equipamento eletrónico • Assunto: transportes entraves técnicos aproximação das legislações					

Optámos por seleccionar textos oriundos da Comissão Económica das Nações Unidas para a Europa (UNECE), em função do papel importante que esta entidade tem em matéria de segurança rodoviária e na implementação de regulamentos especificamente aplicáveis à fabricação de automóveis, no que diz respeito aos padrões de qualidade, segurança e normalização dos diversos dispositivos que os compõem.

*“The UNECE Sustainable Transport Division has been providing secretariat services to the World Forum for Harmonization of Vehicle Regulations (WP.29) for more than 50 years. The World Forum has incorporated into its regulatory framework the technological innovations of vehicles to make them safer and more environmentally sound”.*<sup>24</sup>

No corpo dos documentos recolhidos foi possível encontrar uma hiperligação<sup>25</sup> que remete para o sítio oficial desta entidade, onde se encontra disponível o documento original, sendo que este só se encontra disponível em língua inglesa. O próprio sítio da UNECE só disponibiliza informação para consulta nas línguas francesa, inglesa ou russa, invalidando a nossa utilização do documento original como parte do nosso *corpus*.

### 3.2. Problemas encontrados

Durante o processo de seleção dos textos, houve, em alguns casos, documentos que estavam repetidos. Estes documentos não são na verdade cópias, alguns são versões mais

<sup>24</sup> <http://www.unece.org/trans/main/welcwp29.html>

<sup>25</sup> <http://www.unece.org/trans/main/welcwp29.html>

recentes que sofrem atualizações e servem de referência no seu todo, outros são anexos cujas alterações ou acrescentos substituem partes específicas de um documento maior. Alguns documentos são alvo de várias alterações ao longo dos anos, originando diversas versões do mesmo documento.

Em virtude desta situação, estabeleceu-se que, em caso de repetição, seria descartado o documento cuja data de publicação fosse mais antiga, tendo como referência o ano de 2016.

O regulamento 110 surgiu duas vezes durante a nossa pesquisa. Nesta situação, um dos documentos apresentava a data de publicação de 7/05/2011<sup>26</sup> enquanto o outro tinha sido publicado em 30/6/2015<sup>27</sup>. Foi decidido descartar o documento publicado em 2011 por já existir uma versão do mesmo mais recente, tendo-se agido da mesma forma em todas as ocasiões em que este tipo de conflito surgiu.

Numa fase seguinte foi necessário excluir documentos cujo conteúdo não apresentasse compatibilidade com os nossos objetivos. Saliente-se ainda que não nos foi possível determinar em que língua os textos foram originalmente redigidos e a partir de qual foram traduzidos.

Em função do processo de filtragem ter ficado definido pelos termos introduzidos nos campos de pesquisa da base de dados, refira-se que houve lugar à escolha de documentos cujo cabeçalho não apresentava nenhum dos termos por nós definidos, mas pudemos confirmar a sua presença no corpo do texto dos documentos, tendo, após esta filtragem, restado 23 documentos destinados a fazer parte do nosso *corpus*.

---

<sup>26</sup> [http://eur-lex.europa.eu/legal-content/PT/TXT/?qid=1459291872325&uri=CELEX:42015X0630\(01\)](http://eur-lex.europa.eu/legal-content/PT/TXT/?qid=1459291872325&uri=CELEX:42015X0630(01))

<sup>27</sup> [http://eur-lex.europa.eu/legal-content/PT/TXT/?qid=1459291872325&uri=CELEX:42011X0507\(01\)](http://eur-lex.europa.eu/legal-content/PT/TXT/?qid=1459291872325&uri=CELEX:42011X0507(01))

#### 4. Tratamento semiautomático do *corpus*: ferramentas de análise de *corpora*

Uma vez concluída a tarefa de constituição do *corpus*, segue-se a tarefa de o analisar, recorrendo para tal, a um concordanceiro. Algumas das operações que podemos realizar com este tipo de ferramenta são, por exemplo, processar grandes quantidades de texto, quantificar ocorrências de formas, identificar irregularidades, extrair candidatos a termos e identificar colocações.

Pode encontrar-se uma grande variedade de ferramentas disponíveis atualmente para realizar análises linguísticas. No sítio da Termcoord<sup>28</sup> encontram-se listados alguns destes programas com informação sobre os mesmos.

Na página da Termcoord <sup>29</sup> intitulada “Term extraction tools”, podemos encontrar uma amostra dos principais programas de extração de terminologia existentes. Aqui eles encontram-se organizados em programas não gratuitos e gratuitos.

No que se refere aos últimos, encontramos referidos, entre outros, os programas “TermoStat Web 3.0”, “Lexterm” e “TermMine”, podendo ser necessário efetuar um registo para serem utilizados, mas não comportam qualquer custo para os utilizadores.

No que diz respeito aos programas não gratuitos, destacamos o “SynchroTerm”, “SDL MultiTerm Extract” e o “AlchemyAPI”. No caso de alguns destes programas, pode estar disponível uma versão experimental que possibilita testar as principais funcionalidades do programa, sendo estas desbloqueadas na totalidade em função da sua compra.

Ainda na mesma página, onde se encontram os programas atrás referidos, podemos encontrar o programa de extração de concordâncias que escolhemos para realizar a análise ao nosso *corpus*, o AntConc.

---

<sup>28</sup> <http://termcoord.eu/>

<sup>29</sup> <http://termcoord.eu/discover/free-term-extractors/term-extraction-tools/>

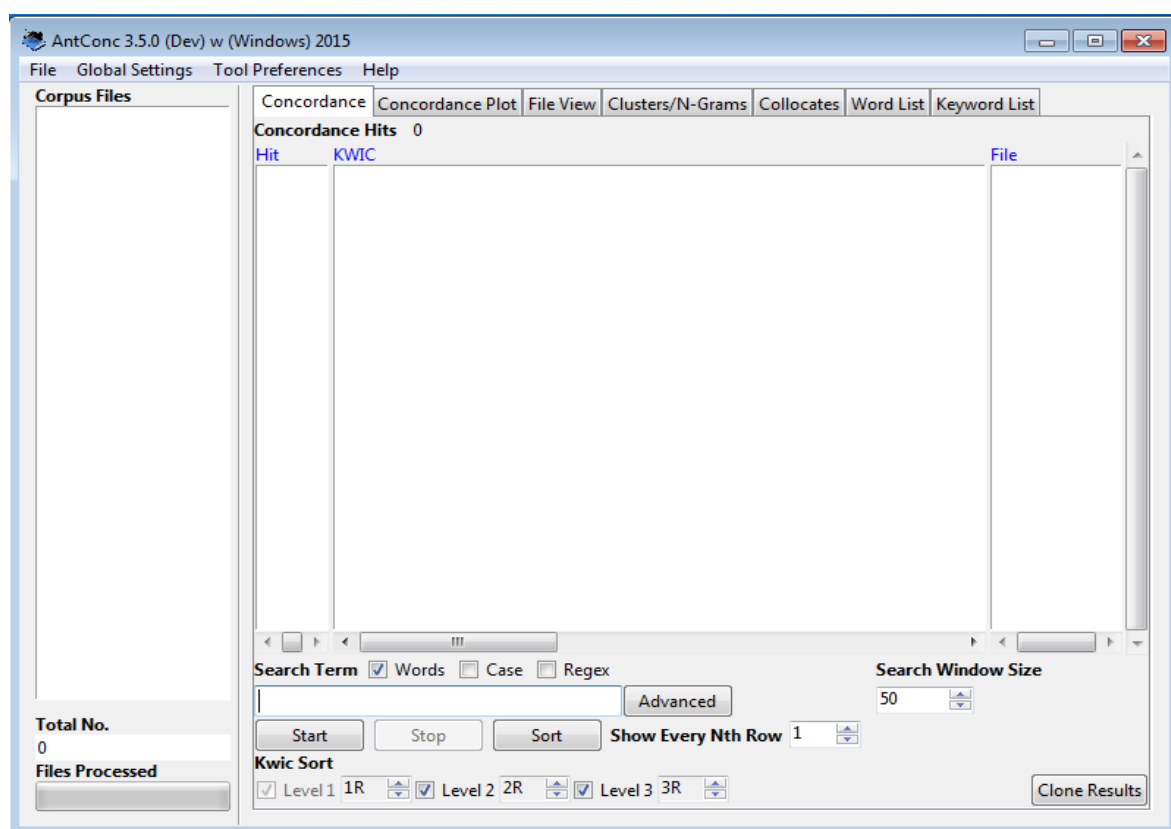
## 4.1. O AntConc: principais características

O AntConc é um programa que está disponível para descarregar, de forma livre e gratuita, sendo de fácil instalação e compatível com os principais sistemas operativos existentes. Na página onde se encontra<sup>30</sup> está disponível um ficheiro em formato PDF, intitulado de HELP, que contém uma descrição acerca das principais funcionalidades e características do concordanceiro, funcionando como um manual de instruções.

Com base neste “manual de instruções”, em inglês, iniciamos uma breve apresentação do AntConc, com ênfase nas funcionalidades mais utilizadas e indispensáveis a uma tarefa de análise de *corpora*.

Para começar a utilizar o AntConc, pressionamos um ícone com a imagem de uma formiga, para abrir o ambiente de trabalho:

**Figura 7 – Ambiente de Trabalho do AntConc**



<sup>30</sup> <http://www.laurenceanthony.net/software.html>

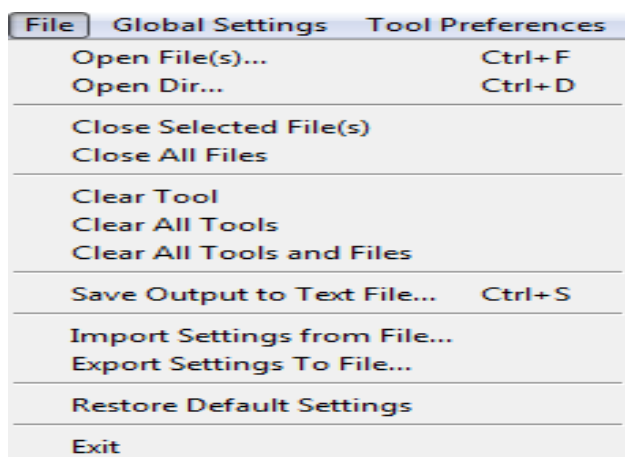
Como podemos verificar atrás, o ambiente de trabalho divide-se em três áreas de trabalho distintas, que designamos por: Menu principal; Menu de ferramentas; Menu de busca, que passamos a descrever de seguida, começando pelo Menu principal:

**Figura 8 - Menu principal**



O Menu principal disponibiliza quatro opções de personalização para a utilização do AntConc. Este concordanceiro funciona com ficheiros em formato txt, sendo necessário a sua conversão para este formato antes de iniciar a seleção dos textos que é feita acedendo ao botão de seleção múltipla “File” do Menu principal que abre a seguinte janela:

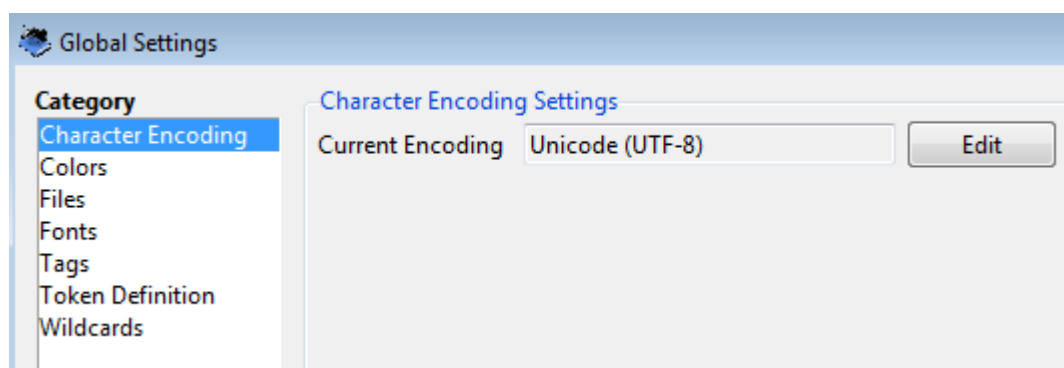
**Figura 9 - File**



O botão de seleção múltipla “File”, representado atrás, abre uma janela que disponibiliza diversas formas de acesso aos ficheiros que se pretende analisar, permitindo uma organização mais conveniente. É aqui que encontramos as opções para importar e exportar os ficheiros, assim como a opção de reinício do programa, colocando os parâmetros de seleção de volta ao seu estado inicial.

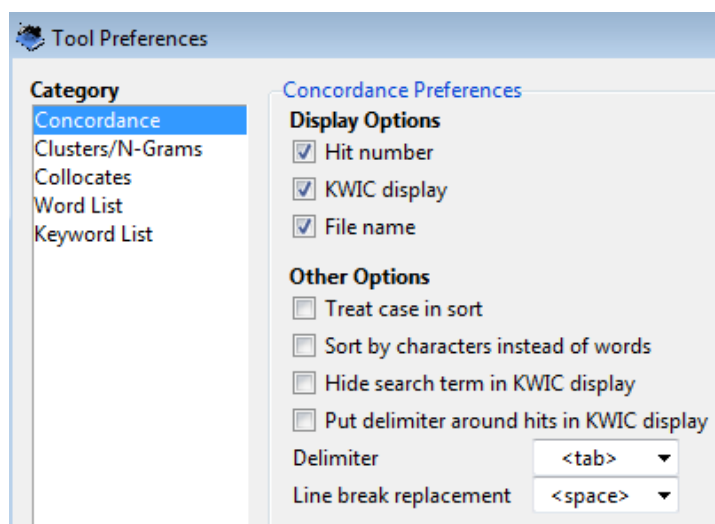
A seguir ao botão seleção múltipla “File” no Menu Principal, encontramos o botão de seleção múltipla “Global Settings”. Este possibilita aceder a várias opções de personalização dos resultados produzidos pelas ferramentas. Um dos campos, o “Character Encoding”, permite escolher a codificação dos caracteres compatível com a língua que se utiliza, sendo recomendado a utilização do padrão internacional “Unicode UTF-8”, como podemos verificar na figura seguinte:

**Figura 10 – Global Settings**



Ainda no Menu principal podemos encontrar o botão de seleção múltipla “Tool Preferences”. Este permite aceder a opções relativas à forma como as ferramentas estão preparadas para produzir os resultados. Um exemplo é a opção de esconder a janela referente ao nome dos ficheiros (File name) que aparece no campo dos resultados produzidos ao utilizar a ferramenta “Concordance”<sup>31</sup>. Ainda no mesmo campo aberto pelo botão de seleção múltipla “Tool Preferences” podemos, ao usar a ferramenta “Clusters/N-Grams”, optar por distinguir entre letras maiúsculas e minúsculas, podendo verificar-se diferenças nos resultados obtidos.

**Figura 11 – Tool Preferences**



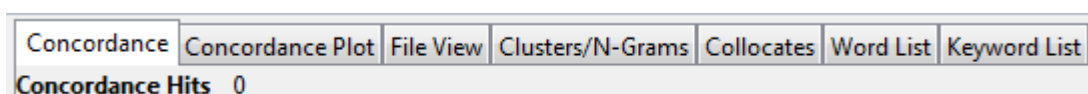
---

<sup>31</sup> Figura 12

Para finalizar a descrição das ferramentas do Menu principal, passamos ao botão seleção múltipla, “Help”, tratando-se este apenas de uma opção que garante acesso a um ficheiro com algumas informações sobre o programa AntConc, como o número e a data da versão do programa que se está a utilizar.

Continuando a nossa descrição das características do AntConc, passamos de seguida ao Menu de ferramentas, onde são disponibilizadas diversas funções de pesquisa e tratamento de *corpora*.

**Figura 12 - Menu de ferramentas**



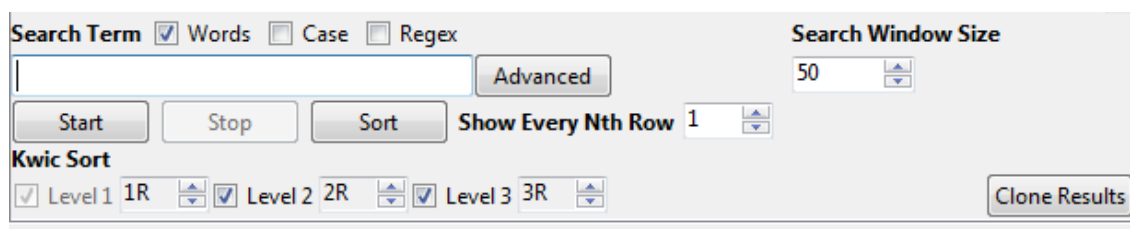
Estas funções são bastante úteis e, através da sua utilização, podemos conduzir uma análise exhaustiva das formas contidas no *corpus*. Para melhor descrever as suas principais características escolhemos dispô-las na seguinte tabela:

**Tabela 3 – Opções do Menu de ferramentas**

Ferramenta	Resultado	Descrição
Concordance	Concordance Hits	A seleção desta ferramenta permite observar o comportamento das formas em contexto
Concordance Plot	Concordance Hits/Total Plots	A seleção desta ferramenta mostra o posicionamento das formas no texto, com um gráfico semelhante a um código de barras
File View	File view Hits/File	A seleção desta ferramenta permite observar uma forma de um texto específico do <i>corpus</i>
Clusters/N-Grams	Total No.of Cluster Types / Total No.of Cluster Tokens	A seleção desta ferramenta permite observar as expressões mais frequentes de um <i>corpus</i>
Collocates	Total No.of collocate Types/ Total No.of collocate Tokens	A seleção desta ferramenta permite observar as colocações de uma forma selecionado
Word List	Word Types/Word Tokens/Search Hits	A seleção desta ferramenta produz uma lista ordenada de todas as formas presentes no <i>corpus</i>
Keyword List:	Types Before Cut/Types After Cut/Search Hits	A seleção desta ferramenta permite verificar a ocorrência das formas pouco usuais no <i>corpus</i> , comparando com um <i>corpus</i> de referência

O Menu de busca permite o estabelecimento de parâmetros específicos de busca em função das necessidades de análise do utilizador.

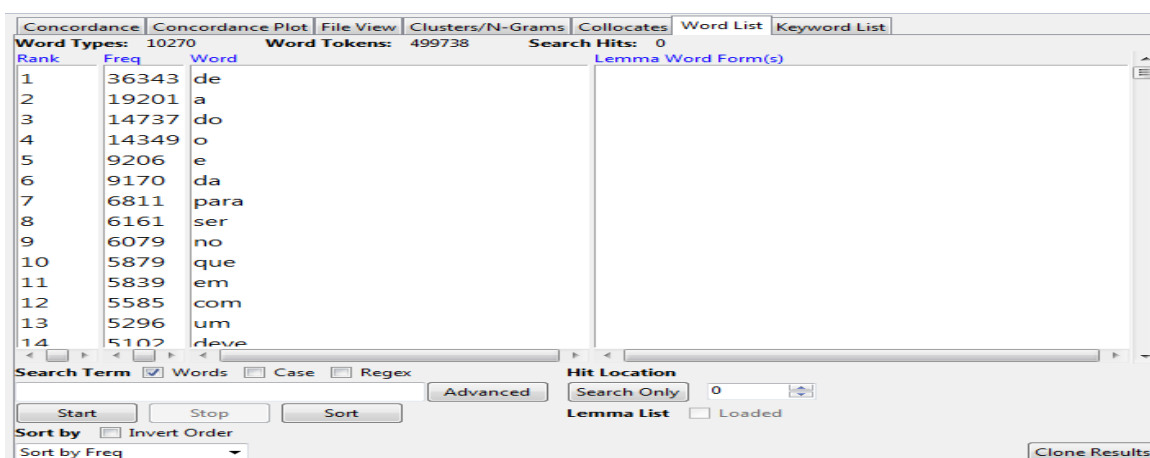
**Figura 13 - Menu de busca**



No Menu de busca atrás representado podemos ver uma caixa de pesquisa, onde se pode introduzir uma ou várias formas e realizar uma pesquisa nos ficheiros de texto. Podemos também seleccionar uma forma directamente do campo de resultados da pesquisa<sup>32</sup>, que por sua vez irá surgir na caixa de pesquisa do Menu de opções de busca. Uma vez aqui podemos alterar os parâmetros da busca, como por exemplo, alterar a ordem da lista de formas resultante ou, dependendo da ferramenta seleccionada do Menu de ferramentas,<sup>33</sup> optar por dar relevo às formas que ocorrem à direita ou à esquerda da forma destacada na caixa de pesquisa.

Como exemplo veja-se o que acontece quando se realiza um busca de concordância, seleccionando o botão “Start” do Menu de busca de busca<sup>34</sup> e a ferramenta “Word List” do Menu de ferramentas<sup>35</sup>:

**Figura 14 – Lista de formas ordenada pelas frequências mais altas**



<sup>32</sup> Figura 14

<sup>33</sup> Figura 12

<sup>34</sup> Figura 13

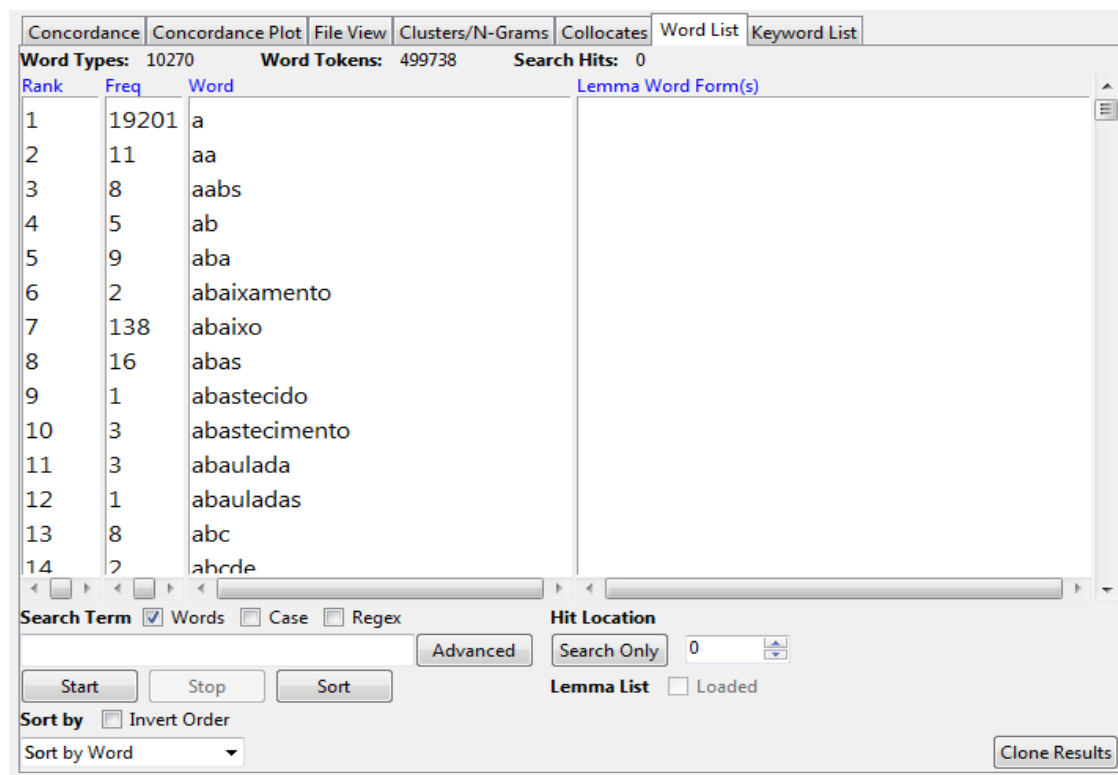
<sup>35</sup> Figura 12



Esta pesquisa produz uma lista de todas as formas encontradas no *corpus*, sendo os resultados apresentados por ordem crescente, da mais frequente para a menos frequente, estando em primeiro lugar a forma [de] com 36343 ocorrências.

Vejamos o que sucede quando alteramos no Menu de busca<sup>36</sup> o parâmetro de “Sort by Freq” para “Sort by Word”:

**Figura 15 – Lista de palavras ordenada alfabeticamente**



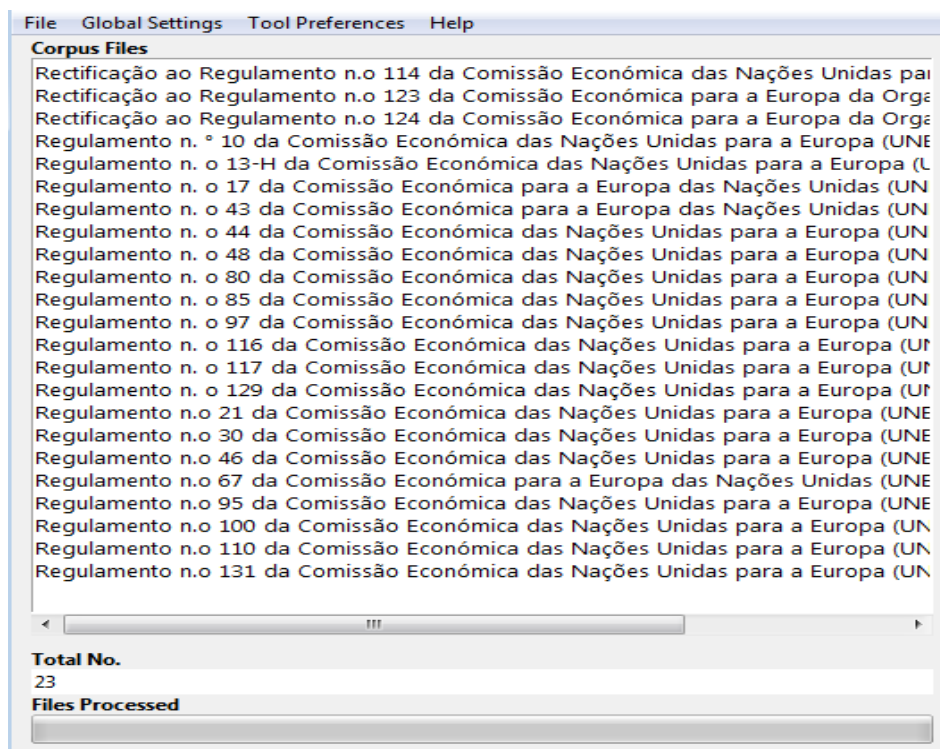
Desta feita, a listagem produzida encontra-se organizada por ordem alfabética e a forma que aparece na primeira posição da lista é [a], que na Figura 14 ocupava a segunda posição.

A personalização dos parâmetros de busca, nos diversos menus, permite analisar o conteúdo de um *corpus* de maneira a satisfazer as necessidades de pesquisa e de investigação.

Só para terminar esta breve introdução às principais funcionalidades do AntConc, mostramos de seguida, a área de trabalho onde se encontram expostos os títulos dos ficheiros de texto seleccionados e disponíveis para análise.

<sup>36</sup>Figura 13

**Figura 16 – Identificação dos ficheiros de texto**



Na figura atrás, podemos verificar que estão selecionados para análise 23 ficheiros de texto, identificados pelo título completo.

A exposição anterior teve como objetivo dar a conhecer as funções principais do AntConc. No capítulo seguinte, recorrendo a este, pretendemos fazer uma aplicação prática das suas funções a um *corpus* de especialidade, com o propósito de estabelecer uma base para uma metodologia de análise de *corpora* com possíveis aplicações ao ensino de língua de especialidade.

## 5. Determinação do campo semântico do domínio

Quando confrontados com a tarefa de identificar termos que pertencem ao domínio da Segurança Automóvel, verificamos que não existe uma metodologia nem critérios sólidos que possibilitem um levantamento de termos de forma a afirmar conclusivamente que os que foram escolhidos fazem parte do domínio.

A nossa dúvida assenta no facto de poder existir um ou mais termos que podem designar um mesmo conceito e, além disso, o mesmo termo poder designar conceitos distintos, pertencentes a domínios também eles diferentes.

A variação existente na designação dos conceitos torna difícil estabelecer os princípios a seguir, por forma a justificar a opção de inclusão de um determinado termo para análise em detrimento de outro.

A abordagem da semântica de *corpus* consiste no estudo da língua a partir de dados recolhidos e reunidos em textos, que atestam a forma como as palavras são usadas em contexto de uso real da língua.

Uma vez que a nossa abordagem consiste em recorrer a um *corpus* para o submeter a uma análise através da aplicação de um concordanceiro, importa determinar qual é a unidade lexical mínima que será alvo de análise. Como tal, começamos por quantificar as unidades que compõem os textos. O que o concordanceiro quantifica, numa primeira fase, são as formas distintas que podem ocorrer num texto, sendo que, por forma se entende, conforme Stubbs (2001, p.26) afirma, “ (...) *the units which occur in actual texts, and in a written text they are strings of letters separated by spaces or punctuation marks.*”. Este autor entende que as formas, dispostas umas após outras, linearmente, são o que define o que é um texto.

A disposição linear é importante, pois de acordo com Stubbs (1991, p.13), “(...) *a text consists of a sequence of word-forms. The sequence is crucial: if we change the sequence we have changed the text.*”

As palavras não têm um sentido fixo, que fica registado perenemente nos dicionários, como Stubbs (1991, p.13) afirma, “*They acquire or change, meaning according to the social and Linguistic contexts in which they are used.*” A ambiguidade que uma palavra isolada pode exibir é desfeita no contexto.

Esta abordagem mostra que o valor semântico pode derivar da combinação entre uma ou mais formas, sendo preciso definir uma outra unidade que dê conta da relação não unívoca entre forma e valor semântico, a saber, unidade lexical.

O termo ‘unidade lexical’ é adotado por vários autores, entre eles, Cruse (1986, p.23) que afirma que as unidades lexicais podem ser classificadas, nos dicionários, de acordo com:

- a forma gráfica e fonológica
- a função gramatical
- o significado

Esta caracterização não é suficiente quando se trata de delimitar as unidades lexicais que ocorrem num texto e Cruse (1986, p.23) aponta três aspetos importantes a considerar:

- “First of all we must delimit the form of a lexical item syntagmatically (...)”
- “Second, having set up syntagmatic units, we shall observe that many of them appear to operate in a variety of grammatical environments (...)”
- “Finally, it is clear that besides having a variety of grammatical uses, a word form may well display a split semantic personality (...)”

No seguimento desta abordagem, Cruse (1986) refere que para uma unidade lexical ser definida sintagmaticamente, deve, em primeiro lugar, ser pelo menos um constituinte semântico<sup>37</sup> e, em segundo, ser pelo menos uma palavra<sup>38</sup>.

Cruse (1986) determina que as unidades lexicais são “ (...) *form-meaning complexes with (relatively) stable and discrete semantic properties which stand in meaning relations*”, sendo as relações em causa, por exemplo, as relações de sinonímia ou antonímia que decorrem da interação sintagmática das unidades lexicais nos contextos.

Como afirma Cruse (1986, p.100), “ (...) *every word in a sentence interacts semantically with every other word, and also with words in neighbouring sentences*”. A uma palavra é associado um sentido, mas esse sentido pode variar em função dos sentidos das

---

<sup>37</sup> Cruse (1986, p.25) define constituinte semântico como, “Any constituent part of a sentence that bears a meaning which combines with the meanings of the other constituents to give the overall meaning of the sentence (...)”, determinando ainda um constituinte semântico como mínimo no caso deste “ (...) cannot be segmented into more elementary semantic constituents (...)”.

<sup>38</sup> Cruse (1986, p.35) estabelece as propriedades de uma palavra como sendo “ (...) *typically the smallest element of a sentence which has positional mobility - that is, the smallest that can be moved around without destroying the grammaticality of the sentence (ignoring any semantic effects)*.”

outras palavras que partilham o contexto, resultando em unidades lexicais com sentido e dimensão distintos das palavras que as compõem.

Também Stubbs (2001, p.7) afirma que “*Any choice of words creates a mini world or universe of discourse, and makes it likely that other words will be co-selected in the same context*” mas refere que o sentido das palavras varia consoante o seu uso em diferentes contextos linguísticos e sociais.

Como tal, este autor afirma que, o vocabulário de uma língua se encontra estruturado por palavras relacionadas entre si através de diferentes tipos de relações lógicas, de semelhança ou diferença e, podem também, estar associadas a um tópico ou campo semântico específico, como se pode constatar pela afirmação: “*Semantic fields are not merely lists of words related by topic: they are also organized by relations amongst these words*” Stubbs (2001, p.36), referindo ainda que “ (...) *we therefore also need concepts to talk about relations between words and the world (...)* ” Stubbs (2001, p.34).

A unidade lexical pode, assim, ser composta por um número diverso de formas, como resultado da diversidade semântica que resulta das combinações sintagmáticas possíveis das formas que a constituem.

Apesar de estabelecidos os princípios básicos do que é uma unidade lexical, precisamos verificar de que forma estes princípios se aplicam no trabalho terminológico.

Em terminologia a unidade mínima de análise é o termo. Este encontra-se definido pela norma ISO 1087-1 (2000, p.6) como, “*verbal designation of a general concept in a specific subject field*”.

L’Homme (2004, p.33) considera que, “*les termes sont des unités lexicales*”, constituindo “*un sous-ensembles du lexique d’une langue*”, sublinhando que a sua principal particularidade é ter um sentido especializado, “*un sens qui peut être mis en rapport avec un domaine de spécialité.*” e a sua definição está dependente da delimitação feita do domínio de especialidade.

Com o termo como unidade mínima de análise estabelecida, importa agora definir os processos através dos quais iremos extrair estas unidades, partindo do *corpus* que reunimos.

## 5.1. Seleção de termos

Em primeiro lugar, como afirmado anteriormente, quantificaremos as formas que ocorrem no *corpus*. Esta tarefa é rápida e é a função mais básica que se pode realizar com um concordanceiro.

Através deste processo é gerada uma lista de formas que ocorrem no *corpus*. Podemos dividir as formas obtidas em dois tipos: as formas que correspondem a palavras funcionais, que desempenham um papel gramatical e são mais esvaziadas de sentido referencial, como é o caso dos determinantes, pronomes, conjunções e certos verbos auxiliares; e as formas que correspondem a palavras de conteúdo, que veiculam sentido referencial, as designadas classes abertas, compostas por nomes, verbos, adjetivos e advérbios.

Na figura seguinte podemos ver os resultados da frequência de algumas formas contidas no nosso *corpus*:

**Figura 17 – Extrato da lista de frequências do *corpus* do domínio da Segurança Automóvel**

Rank	Freq	Word	Lemma Word Form(s)
1	36343	de	
2	19201	a	
3	14737	do	
4	14349	o	
5	9206	e	
6	9170	da	
7	6811	para	
8	6161	ser	
9	6079	no	
10	5879	que	
11	5839	em	
12	5585	com	
13	5296	um	
14	5102	deve	

Como se pode verificar, nas sete primeiras posições as formas que ocorrem correspondem a artigos e preposições (palavras funcionais), aparecendo somente na oitava posição uma forma que corresponde a uma palavra de conteúdo, um verbo<sup>39</sup>.

O concordanceiro permite diferentes métodos de organizar as formas que ocorrem nos textos, mas podemos afirmar que este é um bom ponto de partida e pode ser a base para uma análise de *corpora* para, por exemplo, destacando as formas mais frequentes, realizar uma análise de concordância e verificar o comportamento das formas no *corpus*.

A metodologia que procuramos estabelecer pretende ser uma tentativa de estreitar os processos de seleção de unidades lexicais e adequá-los ao plano específico da terminologia de especialidade.

Nesta abordagem não podemos assumir que, por uma forma ter uma frequência de ocorrência elevada, ela corresponderá a um termo. Sobre o atrás declarado, L'Homme (2004, p.58) afirma que “*Une forme n'apparaissant qu'une seule fois peut être un terme. À l'inverse, des unités linguistiques fréquentes ne sont pas forcément toutes des termes*”. Como tal, é preciso determinar, justificar e extrair as formas que correspondam a termos pertencentes ao domínio da Segurança Automóvel.

De acordo com o atrás estabelecido e apoiando-nos no formulado por L'Homme (2004, p.64) passamos a reconhecer que, qualquer unidade lexical associada ao domínio da Segurança Automóvel terá grande probabilidade de ser um termo do referido domínio.

A nossa análise terá como ponto de partida os termos ‘segurança’ e ‘automóvel’. Propomos realizar uma análise de concordância a ambos os termos por forma a determinar se as nossas expectativas se confirmam, de acordo com os seguintes critérios:

- i. Pesquisar o termo complexo que designa o domínio: ‘segurança automóvel’
- ii. Pesquisar o termo ‘segurança’ que faz parte do termo complexo que designa o domínio
- iii. Pesquisar o termo ‘automóvel’ que faz parte do termo complexo que designa o domínio

---

<sup>39</sup> Partimos do pressuposto de que a forma ‘ser’ corresponde a um verbo, contudo, conforme a nossa proposta, à forma ‘ser’ pode corresponder a uma outra função gramatical, de acordo com o contexto de ocorrência, conforme Bowker e Pearson (2002, p.114) “ (...) *frequency lists show us words out of context, all of the possible interpretations of a particular word will not be immediately apparent.* ”

- iv. Analisar as concordâncias resultantes, com foco acentuado nas formas que ocorrem à esquerda e à direita dos termos pesquisados em ii e iii por forma a isolar unidades lexicais admissíveis a ser consideradas termos pertencentes ao domínio

A escolha das unidades lexicais terá que levar em conta a existência de algum tipo de relação entre os termos de ii e iii e as unidades lexicais candidatas a termos que justifique a escolha, sendo que daremos prioridade às relações semânticas no plano sintagmático.

A análise a realizar no ponto iv, incidirá sobre um contexto composto pelas formas contidas no intervalo entre dois pontos finais, ou seja, um parágrafo.

Conforme estabelecido no ponto i, recorrendo ao concordanceiro, fizemos uma pesquisa pelo termo ‘segurança automóvel’, tendo-se verificado que este não ocorre por nenhuma ocasião em nenhum texto do *corpus* por nós reunido.

Quanto ao determinado no ponto ii, a pesquisa conduz à obtenção de um total de 337 ocorrências do termo ‘segurança’.

De seguida, de acordo com o estabelecido no ponto iii, obtém-se um total de 7 ocorrências para o termo ‘automóvel’.

Avançamos para o ponto iv, no intuito de analisar as concordâncias obtidas nos pontos ii e iii. Decidimos, em função da baixa frequência de ocorrência do termo ‘automóvel’, por contraste com ‘segurança’, centrar a análise exclusivamente neste último, ignorando o estabelecido no ponto iii.

No que respeita ao ponto ii, as linhas de concordância, apresentam o termo pesquisado numa posição central, com seis, sete formas de cada lado. Podemos verificar que a organização dos resultados se encontra, em primeiro, agrupada em função dos textos a que pertencem e, em segundo, pela ordem em que ocorrem nos mesmos. É por esse motivo que, como se pode observar na figura seguinte, temos /cintos de segurança/ a ocorrer na posição 1 e, novamente, na posição 20.



**Figura 18 - Resultado da extração das concordâncias do termo ‘segurança’**

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
<b>Concordance Hits 337</b>						
Hit	KWIC					
1	retenção adicional para além dos cintos de <b>segurança</b> e de outros sistemas de retenção em					
2	retenção adicional para além dos cintos de <b>segurança</b> e de outros sistemas de retenção em					
3	seja susceptível de influenciar o grau de <b>segurança</b> e a resistência mecânica, d) material, e					
4	, sistema dos bancos e dos cintos de <b>segurança</b> e outros sistemas de retenção, d) sistem					
5	de fissuras ou fracturas que afectem a <b>segurança</b> funcional do volante. 6.2.1.3. E n s a					
6	de fissuras ou fracturas que afectem a <b>segurança</b> funcional do volante. O aro é então					
7	a um binário de 22 daNm $\pm$ 0,5 daNm. A <b>segurança</b> funcional do volante não deve ser afect					
8	fissuras ou fracturas que afectem a sua <b>segurança</b> funcional. 6.2.2. Ensaio das prescriçõe					
9	ras informações relevantes; 2.2.2.1 O conceito de <b>segurança</b> , nos termos em que está definido na					
10	qualquer outro processo adaptado às condições de <b>segurança</b> . 2.2.2.2 A marca e o tipo do ou					
11	do fabricante, por força do princípio de <b>segurança</b> estabelecido no ponto 2.2.2.1 supra. 6.2					
12	] $\mu$ = coeficiente de atrito S = coeficiente de <b>segurança</b> 3. O ensaio é realizado com dois valores					
13	normas: Coeficiente de atrito 0,9 Coeficiente de <b>segurança</b> 2,0 Ciclos nominais por minuto O número					
14	carga da roda [N] S = coeficiente de <b>segurança</b> 3. MÉTODO DE ENSAIO E PRESCRIÇÕES: Os en					
15	de rolamento Em linha recta Coeficiente de <b>segurança</b> - S 2,5 2,25 (1) 2,0 Pneumáticos Extraid					
16	= Binário de ensaio [Nm] S = coeficiente de <b>segurança</b> FV = capacidade máxima de carga da roda					
17	com base nos seguintes parâmetros: Coeficiente de <b>segurança</b> - S 1,0 Número mínimo de ciclos a $\pm$ 90 p					
18	ontagem detalhadas: Recomendações e precauções de <b>segurança</b> ao instalar a roda; utilização de compon					
19	de ar e sistemas de retenção de <b>segurança</b> ; c) Funções que, quando perturbadas, pr					
20	..... 53. Cintos de <b>segurança</b> e/ou outros sistemas de retenção: .....					
21	de ar e sistemas de retenção de <b>segurança</b> operacionais, com a almofada de ar do					
22	8. Requisitos especiais a aplicar aos aspectos de <b>segurança</b> dos sistemas complexos de comando eletró					
23	anexo 8 devem ser aplicados aos aspectos de <b>segurança</b> de todos os sistemas complexos de comand					
24	para manutenção e apresentarem características de <b>segurança</b> , pelo menos, iguais às requeridas para a					
25	fonte deve dar a máxima garantia de <b>segurança</b> . 5.2.4.1. Em caso de avaria numa parte					
26	8 REQUISITOS ESPECIAIS A APLICAR AOS ASPETOS DE <b>SEGURANÇA</b> DOS SISTEMAS COMPLEXOS DE COMANDO ELETRÓ					
27	de anomalias no tocante aos aspectos de <b>segurança</b> dos sistemas complexos de comando eletró					

Note-se na figura seguinte os textos a que correspondem as linhas de concordância apresentadas na figura anterior:

**Figura 19 – Identificação dos textos que contêm o termo ‘segurança’**

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
<b>Concordance Hits 337</b>						
Hit	File					
1	Rectificação ao Regulamento n.o 114 da Comissão Económica das Nações Unidas para a Europa (UNECE).txt					
2	Rectificação ao Regulamento n.o 114 da Comissão Económica das Nações Unidas para a Europa (UNECE).txt					
3	Rectificação ao Regulamento n.o 114 da Comissão Económica das Nações Unidas para a Europa (UNECE).txt					
4	Rectificação ao Regulamento n.o 114 da Comissão Económica das Nações Unidas para a Europa (UNECE).txt					
5	Rectificação ao Regulamento n.o 114 da Comissão Económica das Nações Unidas para a Europa (UNECE).txt					
6	Rectificação ao Regulamento n.o 114 da Comissão Económica das Nações Unidas para a Europa (UNECE).txt					
7	Rectificação ao Regulamento n.o 114 da Comissão Económica das Nações Unidas para a Europa (UNECE).txt					
8	Rectificação ao Regulamento n.o 114 da Comissão Económica das Nações Unidas para a Europa (UNECE).txt					
9	Rectificação ao Regulamento n.o 123 da Comissão Económica para a Europa da Organização das Nações Unidas.txt					
10	Rectificação ao Regulamento n.o 123 da Comissão Económica para a Europa da Organização das Nações Unidas.txt					
11	Rectificação ao Regulamento n.o 123 da Comissão Económica para a Europa da Organização das Nações Unidas.txt					
12	Rectificação ao Regulamento n.o 124 da Comissão Económica para a Europa da Organização das Nações Unidas.txt					
13	Rectificação ao Regulamento n.o 124 da Comissão Económica para a Europa da Organização das Nações Unidas.txt					
14	Rectificação ao Regulamento n.o 124 da Comissão Económica para a Europa da Organização das Nações Unidas.txt					
15	Rectificação ao Regulamento n.o 124 da Comissão Económica para a Europa da Organização das Nações Unidas.txt					
16	Rectificação ao Regulamento n.o 124 da Comissão Económica para a Europa da Organização das Nações Unidas.txt					
17	Rectificação ao Regulamento n.o 124 da Comissão Económica para a Europa da Organização das Nações Unidas.txt					
18	Rectificação ao Regulamento n.o 124 da Comissão Económica para a Europa da Organização das Nações Unidas.txt					
19	Regulamento n.º 10 da Comissão Económica das Nações Unidas para a Europa (UNECE).txt					
20	Regulamento n.º 10 da Comissão Económica das Nações Unidas para a Europa (UNECE).txt					
21	Regulamento n.º 10 da Comissão Económica das Nações Unidas para a Europa (UNECE).txt					
22	Regulamento n.º 13-H da Comissão Económica das Nações Unidas para a Europa (UNECE).txt					
23	Regulamento n.º 13-H da Comissão Económica das Nações Unidas para a Europa (UNECE).txt					
24	Regulamento n.º 13-H da Comissão Económica das Nações Unidas para a Europa (UNECE).txt					
25	Regulamento n.º 13-H da Comissão Económica das Nações Unidas para a Europa (UNECE).txt					
26	Regulamento n.º 13-H da Comissão Económica das Nações Unidas para a Europa (UNECE).txt					
27	Regulamento n.º 13-H da Comissão Económica das Nações Unidas para a Europa (UNECE).txt					

A organização dos dados apresentada desta forma não é adequada à nossa pesquisa. Como tal, é necessário ajustar o concordanceiro por forma a listar os resultados de forma mais sintética.

Como o nosso objetivo é dispor os resultados em função das formas que ocorrem à esquerda e à direita de ‘segurança’, começamos por programar o concordanceiro para mostrar os resultados da combinação com as formas à direita.

No Menu de ferramentas selecionamos a ferramenta “Clusters/N-Grams”, ajustando as definições do Menu de busca para selecionar três formas à esquerda de ‘segurança’, estando já esta incluída. Através desta filtragem obtemos uma lista das combinações mais frequentes, organizadas em função do número total de ocorrências, da mais para a menos frequente.

De um total de 74 ocorrências, apresentamos de seguida a lista das dez unidades lexicais mais frequentes que contêm o termo ‘segurança’:

**Figura 20 - lista das dez unidades lexicais mais frequentes associadas à esquerda do termo ‘segurança’**

Concordance Concordance Plot File View Clusters/N-Grams Colli				
Total No. of Cluster Types 74			Total No. of Cluster Tok	
Rank	Freq	Range	Cluster	
1	60	6	cinto de segurança	
2	58	1	vidraças de segurança	
3	30	8	cintos de segurança	
4	28	1	vidraça de segurança	
5	13	3	conceito de segurança	
6	13	6	condições de segurança	
7	9	1	matéria de segurança	
8	7	2	aspetos de segurança	
9	6	2	bloqueamento de segurança	
10	6	1	coeficiente de segurança	

Na posição 1 temos a unidade lexical /cinto de segurança/, com um total de 60 ocorrências. A estas devemos acrescentar o número de ocorrências registado pela unidade lexical, /cintos de segurança/ (plural) na terceira posição com um total de 30 ocorrências.

O nosso conhecimento acerca do domínio não pode ser considerado ao nível do de um especialista, mas podemos afirmar, de acordo com o nosso conhecimento pessoal, que ‘cinto de segurança’ é um termo que designa um objeto que faz parte de um veículo automóvel e cujo propósito é aumentar a segurança do condutor do mesmo.

Uma vez que temos o *corpus* à nossa disposição, realizamos uma pesquisa isolada ao termo ‘cinto de segurança’, mas alargamos o escopo da nossa pesquisa, incluindo um número maior de formas a seguir a ‘cinto de segurança’. Este alargamento permite verificar que, ao acrescentar outras formas, obtemos unidades lexicais distintas, ainda que associadas ao termo ‘cinto de segurança’, que conseguimos determinar em função do contexto que é criado com o citado alargamento.

Expomos de seguida os 23 resultados obtidos para um grupo constituído por cinco formas, com base no termo ‘cinto de segurança’:

**Figura 21 - lista das unidades lexicais criadas a partir do termo ‘cinto de segurança’**

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Ke
Total No. of Cluster Types		23	Total No. of Cluster Tokens		60	
Rank	Freq	Range	Cluster			
1	34	2	cinto de segurança para adultos			
2	2	1	cinto de segurança a utilizar			
3	2	2	cinto de segurança colocado no			
4	2	2	cinto de segurança de dois			
5	2	2	cinto de segurança equipado com			
6	1	1	cinto de segurança apertado; 3.4.1.2. com			
7	1	1	cinto de segurança de três			
8	1	1	cinto de segurança montado e			
9	1	1	cinto de segurança normalizado apropriado			
10	1	1	cinto de segurança ou sistema			
11	1	1	cinto de segurança para efeitos			
12	1	1	cinto de segurança utilizado deve			
13	1	1	cinto de segurança). 2.5. «isofix», sistema			
14	1	1	cinto de segurança-padrão 1. o			
15	1	1	cinto de segurança-padrão anexo			
16	1	1	cinto de segurança-padrão dimensões			
17	1	1	cinto de segurança-padrão figura			
18	1	1	cinto de segurança-padrão não			
19	1	1	cinto de segurança-padrão utilizado			
20	1	1	cinto de segurança. 2.10.1. «dispositivo de			
21	1	1	cinto de segurança. a orientação			
22	1	1	cinto de segurança. devem ser			
23	1	1	cinto de segurança; 1.1.1.1. este requisito			

Do conjunto de resultados anterior, julgamos ser possível destacar três grupos de formas como tendo propriedades de unidade lexical: /cinto de segurança para adultos/; /cinto de segurança-padrão/ e /cinto de segurança normalizado/, sendo este último constituído por apenas quatro formas.

Apresentamos de seguida o resultado das concordâncias das formas atrás destacadas, apresentando alguns contextos retirados a partir dos textos:

1. “Sistema de retenção parcial», um dispositivo, como uma almofada elevadora, que, quando utilizado juntamente com um **cinto de segurança para adultos**,

passado em volta do corpo da criança ou disposto de forma a reter o dispositivo no qual a criança está colocada, constitui um sistema de retenção para crianças completo.”

2. “Para verificar o cumprimento deste requisito, o sistema de retenção para crianças deve ser fixado ao banco de ensaio por meio do **cinto de segurança normalizado** apropriado descrito no anexo 13.”
3. “Considera-se como limite de movimento aceitável da precinta do ombro que a extremidade inferior da parte escapular do **cinto de segurança**-extremidade inferior da parte escapular do **cinto de segurança-padrão** não se encontre abaixo do cotovelo do manequim no ponto máximo de deslocação deste último”

Com base nos exemplos atrás, podemos observar que as unidades lexicais destacadas, formadas a partir de ‘cinto de segurança’, se associam nos contextos respetivos, com unidades lexicais que nos levam a considerar a hipótese de: /cinto de segurança para adultos;/ /cinto de segurança-padrão/ e /cinto de segurança normalizado/ se referirem a tipos distintos de sistemas de retenção.

No seguimento da nossa hipótese, pesquisamos separadamente ‘cintos de segurança’, e ‘cinto de segurança’ para comprovar que não se regista diferença no valor semântico em relação aos valores de plural e singular, respetivamente e, para verificar se estes termos também têm alguma associação a sistemas de retenção:

4. “**Cintos de segurança** e/ou outros sistemas de retenção:”
5. “1.1. A finalidade dos ensaios é determinar: 1.1.1. se o(s) ocupante(s) do banco é(são) adequadamente retidos pelo banco(s) situado(s) à sua frente e/ou por um **cinto de segurança**;”

Não foi possível obter, a partir do *corpus*, um contexto definatório em relação a ‘cinto de segurança’, mas a análise dos exemplos atrás leva-nos a afirmar que se trata de um sistema de retenção.

Apesar de a nossa metodologia pretender estabelecer maneiras de isolar unidades lexicais candidatas a termos, a partir dos textos, não podemos deixar de constatar que a partir dos textos podem ser identificadas relações paradigmáticas, especificamente no caso do exemplo 4, onde se pode perceber uma relação de subordinação da unidade lexical /cintos de segurança/ em relação à unidade lexical /sistemas de retenção/, o seu hiperónimo.

Voltamos agora a nossa atenção para as formas que ocorrem à direita, de ‘segurança’, verificando-se um total de 203 ocorrências, das quais destacamos de seguida as dez mais frequentes:

**Figura 22 - lista das dez unidades lexicais mais frequentes associadas à direita do termo ‘segurança’**

Concordance	Concordance Plot	File View	Clusters/N-Grams	Colloc
Total No. of Cluster Types		203	Total No. of Cluster Tokens	
Rank	Freq	Range	Cluster	
1	40	2	segurança para adultos	
2	9	2	segurança de funcionamento	
3	9	1	segurança nas mesmas	
4	7	2	segurança dos sistemas	
5	6	2	segurança contra a	
6	6	6	segurança e a	
7	5	1	segurança do sraee	
8	5	1	segurança revestidas de	
9	4	3	segurança do funcionamento	
10	4	1	segurança e o	

Como podemos verificar, em primeiro lugar, com um total de 40 ocorrências surge a unidade lexical /segurança para adultos/, que apesar de muito frequente, não parece ter condições para ser um possível termo.

A nossa dúvida tem por base a análise anterior, em que nos concentrámos nas formas que precediam ‘segurança’. Neste caso foi possível destacar ‘cinto de segurança’, não só por ser muito frequente, mas por ser possível estabelecer uma ligação ao domínio, por força do nosso conhecimento sobre o objeto a que ‘cinto de segurança’ se refere.

Ainda na mesma análise, foi possível verificar que, das 60 ocorrências de ‘cinto de segurança’, em 34 ocasiões esta forma encontrava-se seguida da combinação [para adultos], como se pode verificar na figura seguinte:

**Figura 23 – Resultado das concordâncias das formas [para adultos] precedidas de ‘cinto de segurança’**

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Concordance Hits 34						
Hit	KWIC					
1	levadora, que, quando utilizado juntamente com um <a href="#">cinto de segurança para adultos</a> , passado em volta					
2	firme que pode ser utilizada com um <a href="#">cinto de segurança para adultos</a> . 2.2. «Cadeira de					
3	que sustém a precinta do ombro do <a href="#">cinto de segurança para adultos</a> numa posição adapt					
4	2.23.3. O tipo e dimensões da fixação do <a href="#">cinto de segurança para adultos</a> ao banco, da					
5	duas secções de precinta de um mesmo <a href="#">cinto de segurança para adultos</a> . Esses dispositivo					
6	do cinto e a secção diagonal do <a href="#">cinto de segurança para adultos</a> . O termo abrange					
7	subabdominal do cinto, quando for utilizado um <a href="#">cinto de segurança para adultos</a> para susten direct					
8	tensão aplicada na parte subabdominal de um <a href="#">cinto de segurança para adultos</a> , quando este for					
9	cinta do tirante superior ISOFIX. 2.37. «Guia do <a href="#">cinto de segurança para adultos</a> », um dispositivo d					
10	guiamento por onde passa a precinta do <a href="#">cinto de segurança para adultos</a> e que permite					
11	de retenção para crianças, for utilizado um <a href="#">cinto de segurança para adultos</a> homologado, o pedi					
12	de homologação deve indicar a categoria de <a href="#">cinto de segurança para adultos</a> a utilizar, por					
13	a ser utilizado em combinação com um <a href="#">cinto de segurança para adultos</a> , deve estar perman					
14	de retenção for mantido no lugar pelo <a href="#">cinto de segurança para adultos</a> , as formas de					
15	sistema de retenção para crianças e o <a href="#">cinto de segurança para adultos</a> devem estar marcad					
16	» e «restrito», apenas por meio de um <a href="#">cinto de segurança para adultos</a> (com ou sem					
17	levadora tem de ser retida com um <a href="#">cinto de segurança para adultos</a> , em conformidade c					
18	sistema de retenção para crianças e o <a href="#">cinto de segurança para adultos</a> não deve distar					
19	no presente número. 6.1.9. Se for necessário um <a href="#">cinto de segurança para adultos</a> para fixar um					
20	das precintas e da lingueta de um <a href="#">cinto de segurança para adultos</a> através dos pontos					
21	mecânica, nos quais tanto a altura do <a href="#">cinto de segurança para adultos</a> como do arnês					
22	destinar a ser utilizado juntamente com um <a href="#">cinto de segurança para adultos</a> , as partes em					
23	loqueamento não deve prejudicar a durabilidade do <a href="#">cinto de segurança para adultos</a> , e deve ser					
24	o esforço seja exercido directamente através do <a href="#">cinto de segurança para adultos</a> sobre a fixação					
25	são utilizáveis. 15.2.4. Se o sistema exigir um <a href="#">cinto de segurança para adultos</a> , a seguinte frase					
26	o dispositivo for utilizado juntamente com um <a href="#">cinto de segurança para adultos</a> , a seguinte frase:					
27	insatisfatória da fivela de fecho de um <a href="#">cinto de segurança para adultos</a> em relação aos					

Verificámos ainda 6 ocorrências das formas [para adultos], precedidas de ‘cintos de segurança’ (plural) como se pode constatar na figura seguinte:

**Figura 23 – Resultado das concordâncias das formas [para adultos] precedida de ‘cintos de segurança’**

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Concordance Hits 6						
Hit	KWIC					
1	quais já estejam fixados um ou mais <a href="#">cintos de segurança para adultos</a> , o serviço técn					
2	fixações dos bancos do veículo e dos <a href="#">cintos de segurança para adultos</a> ou quaisquer outr					
3	e utilizem precintas suplementares ligadas a dois <a href="#">cintos de segurança para adultos</a> , em que o					
4	tem de trabalhar correctamente mesmo quando os <a href="#">cintos de segurança para adultos</a> estiverem desbloq					
5	, o sistema de retenção é ligado aos <a href="#">cintos de segurança para adultos</a> com extremidades					
6	fixações dos bancos do veículo e dos <a href="#">cintos de segurança para adultos</a> ou quaisquer outr					

A análise realizada em função da posição do termo ‘segurança’, em relação às formas que com ele ocorrem, apresenta resultados distintos. Quando nos concentrámos nas formas que ocorrem à esquerda de ‘segurança’, foi possível, como já referimos, destacar unidades

lexicais que apresentavam grande probabilidade de se constituírem como termos. No caso das formas que ocorrem à direita, só após consultar os contextos de ocorrência correspondentes às concordâncias foi possível estabelecer, por exemplo, que /segurança para adultos/, era parte de uma unidade maior, a saber, /cinto de segurança para adultos/, cujo valor de possível termo já havíamos estabelecido anteriormente.

Esta abordagem não teve o mesmo sucesso quando a aplicámos às formas que se seguiam na lista de frequências disposta na Figura 22 mais atrás. Veja-se o resultado das concordâncias para a unidade lexical /segurança de funcionamento/:

**Figura 25 - Resultado das concordâncias da unidade lexical /segurança de funcionamento/**

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Concordance Hits 9						
Hit	KWIC					
1	minuto, nos estados ativado e desativado. 7.2.3. <a href="#">Segurança de funcionamento</a> depois dos ensaios de r					
2	homologação da unidade técnica autónoma. 7.2.4. <a href="#">Segurança de funcionamento</a> depois do ensaio de con					
3	homologação da unidade técnica autónoma. 7.2.8. <a href="#">Segurança de funcionamento</a> depois do ensaio de vib					
4	amento seja fiável (ver pontos 31.8 e 33). 32.3. <a href="#">Segurança de funcionamento</a> Deve garantir-se que n					
5	da sua alimentação elétrica, as condições de <a href="#">segurança de funcionamento</a> do veículo não podem se					
6	igual a (24 ± 0,2) V durante um minuto. 6.4.2.3. <a href="#">Segurança de funcionamento</a> depois dos ensaios de r					
7	para a homologação da unidade técnica. 6.4.2.4. <a href="#">Segurança de funcionamento</a> após o ensaio de conden					
8	a a homologação de uma unidade técnica. 6.4.2.8. <a href="#">Segurança de funcionamento</a> depois do ensaio de vib					
9	do veículo (ver os n. os 8.2.8 e 8.4). 8.3.3. <a href="#">Segurança de funcionamento</a> Deve garantir-se que n					

Apenas pela observação das linhas de concordância atrás expostas, não fomos capazes de determinar uma unidade lexical com possibilidade de se tornar um termo do domínio, no entanto, considerámos, ainda que com algumas reservas, após consultar o texto correspondente, clicando no campo identificado como “File View” do menu de ferramentas, que /condições de segurança de funcionamento do veículo/ representada atrás na linha 5, tinha condições para ser uma unidade lexical. A nossa dúvida em relação à sua inclusão como termo pertencente ao domínio da Segurança Automóvel não se desfez.

A hipótese que colocamos é que estas unidades fazem parte de um contexto maior, talvez relacionadas anaforicamente com outras unidades lexicais, mas não nos é possível determinar se são termos sem o recurso a um especialista. Além disso fica por saber qual seria ao resultado se o termo pesquisado fosse outro que não ‘segurança’. Não podemos afastar a possibilidade de a dificuldade encontrada em determinar unidades lexicais com probabilidade de se tornarem termos, seja devida às particularidades do termo ‘segurança’.



Assim, a partir da análise de concordâncias e dos correspondentes contextos, propomos a seguinte lista de prováveis termos, em função da sua associação ao termo ‘segurança’:

- Cinto de segurança;
- Cinto de segurança para adultos;
- Cinto de segurança-padrão;
- Cinto de segurança normalizado;

Para sumarizar, a nossa metodologia consiste em determinar um termo que pertença ao domínio em estudo, a partir do qual se analisa as formas que ocorrem à esquerda ou direita, estabelecendo para tal, um intervalo que contenha formas suficientes.

O intervalo e o número de formas nele contido não necessita de ser rígido, pois como declarámos anteriormente, são raros os casos de contextos definitórios. Como tal achamos necessário não limitar a observação da forma à linha de concordância produzida automaticamente pelo concordanceiro, ainda que a extensão da linha possa ser personalizada, deve-se consultar o ficheiro do texto onde a forma ocorre, pois permite observar outros contextos adjuntos à forma em análise.

Consideramos que esta abordagem pode produzir resultados no caso de se pretender isolar unidades lexicais candidatas a termos. O processo deve ter como base um termo previamente estabelecido como tal. No nosso caso partimos do termo que designa o domínio, mas o ideal será trabalhar em conjunto com um especialista que possa validar as unidades base como termos do domínio. No entanto, um terminólogo ou investigador que analise *corpora* tem desta forma uma metodologia que dê conta de candidatos a termos, sobre os quais pode questionar um especialista acerca da sua validade.

A metodologia atrás estabelecida consiste numa prática que é aplicada em função de objetivos mais gerais, isolar unidades lexicais e tentar determinar a sua propensão a tornarem-se termos de um domínio. No capítulo seguinte iremos apresentar um outro tipo de metodologia que consideramos ser mais específica, por se focar numa única forma [entende-se] que servirá de base à nossa análise.



## 5.2. Análise de uma forma pré-selecionada [entende-se]

Neste capítulo pretende-se, à semelhança do anterior, recorrer ao concordanceiro para levar a cabo uma nova análise do *corpus*. No anterior capítulo, a metodologia consistiu numa análise mais geral. Agora, é nossa intenção partir de uma forma específica, a saber a forma [entende-se] para verificar qual o seu comportamento específico no *corpus* em análise.

A opção pela forma [entende-se] foi tomada em virtude de esta ter surgido durante a fase anterior de análise, inserida em contextos durante a pesquisa de concordâncias. No caso, a pesquisa foi referente ao termo ‘segurança’, que surge enquadrado no contexto “conceito de segurança”. Esta combinatória é precedida da forma [entende-se] e conquistou a nossa atenção, em virtude de evidenciar características comuns a outros termos, que surgem em contextos do tipo definatório e que apresentam uma estrutura específica, no que a definições de tipo contextuais diz respeito.

A análise foi feita a partir da forma [entende-se], com recurso ao concordanceiro, de acordo com os seguintes critérios:

- i. realizar uma pesquisa com o objetivo de determinar a frequência da forma [entende-se] no nosso *corpus*;
- ii. identificar quais as formas que ocorrem à direita da forma [entende-se];
- iii. determinar a categoria gramatical correspondente às formas encontradas no ponto ii;
- iv. analisar as concordâncias resultantes, assim como os contextos correspondentes.

Quanto ao atrás estabelecido, importa acrescentar que, apesar de optarmos por limitar a nossa análise às formas que ocorrem à direita de [entende-se], esta opção não invalida, quando necessário, a análise dos termos que a precedem.

Determinar a categoria gramatical das formas encontradas será feita em função do anteriormente estabelecido quanto ao que se entende por palavras de conteúdo e palavras funcionais.

Estabelecemos que a análise a realizar no ponto iv não se limitará a um número fixo de formas, incidirá sobre as formas contidas no intervalo entre dois pontos finais, ou seja, um parágrafo.

No que respeita ao estabelecido no ponto i, verificamos que a forma [entende-se] ocorre por 96 ocasiões no *corpus*, como se pode observar na figura seguinte:

**Figura 26 - resultado da extração das concordâncias da forma [entende-se]**

Concordance		Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Concordance Hits 96							
Hit	KWIC						
1	DEFINIÇÕES 2.1. Por «sistema de almofada de ar», entende-se um grupo de componentes que, uma						
2	éctricos (se existir). 2.2. Por «almofada de ar», entende-se um material flexível, formando um volum						
3	ocupante. 2.3. Por «módulo de almofada de ar», entende-se o menor subconjunto que compreende a						
4	com um módulo de almofada de ar), entende-se um volante destinado a modificar o						
5	«sistema de almofada de ar de substituição», entende-se um sistema de almofada de ar						
6	colisão lateral. 2.7. Por «instante de abertura», entende-se o momento em que, na eventualidade						
7	«unidade de controlo ou unidade de abertura», entende-se um subconjunto compreendendo todos os c						
8	almofada de ar. 2.9. Por «feixe de fios», entende-se o conjunto dos condutores e cabos						
9	«massa sem carga em ordem de marcha», entende-se a massa do veículo em ordem						
10	sistema de almofada de ar de substituição», entende-se os módulos de almofada de ar						
11	com um módulo de almofada de ar», entende-se os volantes que não difiram entre						
12	substituição diferente do instalado no volante», entende-se os sistemas de almofada de ar						
13	fixação do módulo. 2.14. Por «tipo de veículo», entende-se uma categoria de veículos de comando						
14	DE CRUZAMENTO (1) Para efeitos do presente anexo, entende-se por: «acima», colocado por cima, segund						
15	. Para efeitos do disposto no presente anexo, entende-se por: a) «Sistema completo», o lado						
16	de um lote homogéneo. Por lote homogéneo entende-se um conjunto de sistemas do mesmo						
17	DEFINIÇÕES Para efeitos do presente regulamento, entende-se por: 2.1. «Roda», um elemento rotativo						
18	DEFINIÇÕES Para efeitos do presente regulamento, entende-se por: 2.1. «Compatibilidade eletromagné						
19	DEFINIÇÕES Para efeitos do presente regulamento, entende-se por: 2.1. «Homologação de um veículo»,						
20	ÕES 2.1. Por «sistema de travagem antibloqueio», entende-se um elemento de um sistema de						
21	veículo, durante a travagem. 2.2. Por «sensor», entende-se um componente concebido para detetar e						
22	es dinâmicas do veículo. 2.3. Por «controlador», entende-se um componente concebido para avaliar as						
23	tir um sinal ao modulador. 2.4. Por «modulador», entende-se um componente concebido para variar a(						
24	rolador. 2.5. Por «roda diretamente controlada», entende-se uma roda cuja força de travagem						
25	( 1 ). 2.6. Por «roda indiretamente controlada», entende-se uma roda cuja força de travagem						
26	( 1 ). 2.7. Por «execução de ciclos completos», entende-se a modulação repetida da força de						
27	. 2. DEFINIÇÕES Para efeitos do presente anexo, entende-se por: 2.1. «Conceito de segurança», uma						

Como se pode observar, a forma [entende-se] é a forma polo que apresenta seis a sete formas de cada lado, em cada linha. Voltamos a lembrar que a organização dos resultados se encontra agrupada em função da posição dos textos a que pertencem e, também, pela ordem em que ocorrem nos mesmos.

### 5.2.1. Análise da combinação das formas ‘entende-se um’

Passamos, de acordo com o estabelecido no ponto ii, à extração das formas que ocorrem à direita de [entende-se]. Após a sua inserção na caixa de pesquisa do Menu de busca, pressionamos o botão “Start” e obtemos os seguintes resultados:

**Figura 27 – lista das dez unidades lexicais mais frequentes associadas à direita da forma [entende-se]**

Concordance	Concordance Plot	File View	Clusters/N-Grams
Total No. of Cluster Types		12	Total No. of Cluster
Rank	Freq	Cluster	
1	33	entende-se um	
2	30	entende-se por	
3	10	entende-se a	
4	6	entende-se o	
5	4	entende-se qualquer	
6	4	entende-se uma	
7	3	entende-se os	
8	2	entende-se que	
9	1	entende-se mangas	
10	1	entende-se reservatórios	
11	1	entende-se, salvo	
12	1	entende-se: a	

A pesquisa mostra que, imediatamente à direita da forma [entende-se], ocorrem 12 tipos distintos de formas:

**Tabela 4 - classificação gramatical formas ocorrentes à direita de [entende-se]**

Nº Ocorrências	Forma à direita de entende-se	Classe gramatical
33	um	Artigo indefinido singular masculino
30	por	Preposição
10	a	Artigo definido singular feminino
6	o	Artigo definido singular masculino
4	qualquer	Quantificador
4	uma	Artigo indefinido singular feminino
3	os	Artigo definido plural masculino
2	que	conjunção
1	mangas	Substantivo plural feminino
1	reservatórios	Substantivo plural masculino
1	salvo	advérbio
1	: a <sup>40</sup>	

Constata-se uma predominância de formas correspondentes a palavras funcionais, artigos e preposições, à direita da forma [entende-se]. No que respeita à classificação gramatical, atribuída a algumas das formas, lembramos que esta classificação é provisória, sendo necessário verificar posteriormente, de acordo com o contexto em que ocorrem.

<sup>40</sup> Esta forma diz respeito a um formalismo utilizado para organizar uma listagem, como por exemplo: a) primeiro; b) segundo.

Numa primeira impressão constatamos que predomina uma combinação da forma [entende-se] com formas que correspondem a palavras funcionais. Encontramos na primeira posição, com um total de 33 ocorrências, o artigo indefinido, “um” e, em segunda posição, a preposição “por”, com um total de 30 ocorrências.

Tomando em consideração a natureza gramatical das formas anteriores, por pertencerem a tipos diferentes de palavras funcionais, quer pela diferença das suas frequências em relação às formas que surgem nas posições seguintes, decidimos limitar a prática da nossa metodologia à análise destas duas formas.

Como já tínhamos referido, a listagem das frequências providenciada pelo concordanceiro, permite quantificar as ocorrências das diferentes formas que ocorrem no *corpus*, sendo no entanto necessário, posteriormente, estreitar o campo das formas em análise.

O primeiro passo que estabelecemos é programar o concordanceiro para mostrar apenas as concordâncias referentes à combinação das formas [entende-se um]. Como podemos ver na figura seguinte, este estreitamento dos resultados continua a não ser suficiente.

**Figura 28- Resultado da extração de concordâncias referente à combinação entre a forma [entende-se] com a forma [um]**

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Concordance Hits 33						
Hit	KWIC					
1	DEFINIÇÕES 2.1. Por «sistema de almofada de ar», entende-se um grupo de componentes que, uma					
2	étricos (se existir). 2.2. Por «almofada de ar», entende-se um material flexível, formando um volum					
3	com um módulo de almofada de ar), entende-se um volante destinado a modificar o					
4	«sistema de almofada de ar de substituição», entende-se um sistema de almofada de ar					
5	«unidade de controlo ou unidade de abertura», entende-se um subconjunto compreendendo todos os c					
6	de um lote homogéneo. Por lote homogéneo entende-se um conjunto de sistemas do mesmo					
7	ÕES 2.1. Por «sistema de travagem antibloqueio», entende-se um elemento de um sistema de					
8	veículo, durante a travagem. 2.2. Por «sensor», entende-se um componente concebido para detetar e					
9	es dinâmicas do veículo. 2.3. Por «controlador», entende-se um componente concebido para avaliar as					
10	tir um sinal ao modulador. 2.4. Por «modulador», entende-se um componente concebido para variar a(					
11	servatório inteiramente em materiais compósitos», entende-se um reservatório feito exclusivamente de					
12	não metálico. 2.3.3. Por «lote de reservatórios», entende-se um máximo de 200 reservatórios do mesmo					
13	1. Por «válvula limitadora do enchimento a 80 %», entende-se um dispositivo que não permite encher					
14	sua capacidade. 2.5.2. Por «indicador de nível», entende-se um dispositivo que permite verificar o					
15	«válvula de descompressão (válvula de descarga)», entende-se um dispositivo que limita a subida					
16	2.5.3.1. Por «dispositivo limitador de pressão», entende-se um dispositivo destinado a proteger o					
17	lecomandada, com válvula de limitação do débito», entende-se um dispositivo que permite estabelecer					
18	fechada; por válvula de limitação do débito entende-se um dispositivo para limitar o fluxo					
19	uropeia L 72/5 2.5.5. Por «bomba de combustível», entende-se um dispositivo para fornecer GPL líquido					
20	bomba de combustível. 2.5.6. Por «multiválvula», entende-se um dispositivo composto por todos ou					
21	.o 2.5.8. 2.5.7. Por «cobertura estanque ao gás», entende-se um dispositivo para proteger os acessór					
22	combustível). 2.5.9. Por «válvula anti-retorno», entende-se um dispositivo que permite o fluxo					
23	impede no sentido oposto. 2.6. Por «vaporizador», entende-se um dispositivo destinado a fazer passar					
24	o estado gasoso. 2.7. Por «regulador de pressão», entende-se um dispositivo destinado a reduzir e					
25	o liquefeitos. 2.8. Por «válvula de interrupção», entende-se um dispositivo destinado a travar o					
26	(ou de descarga) da tubagem de gás», entende-se um dispositivo que impede a pressão					
27	do gás, injetor ou misturador do gás», entende-se um dispositivo que determina a admissão					

O nosso objetivo é a partir das concordâncias identificar e isolar um padrão recorrente. Uma observação inicial dos resultados apresentados na figura anterior permite

verificar que há formas que surgem de forma recorrente a seguir a [entende-se um], podendo, no entanto, esta percepção ser enganadora. Em função da necessidade de confirmação de um possível padrão, configuramos o concordanceiro para mostrar apenas os resultados para as combinações entre [entende-se um] e a forma seguinte:

**Figura 29 - lista das dez unidades lexicais mais frequentes associadas à direita da forma [entende-se um]**

Concordance		Concordance Plot	File View	Clusters/N-Grams	Collo
Total No. of Cluster Types		12			
		Total No. of Cluster Tokens			
Rank	Freq	Range	Cluster		
1	20	1	entende-se um dispositivo		
2	3	1	entende-se um componente		
3	1	1	entende-se um conjunto		
4	1	1	entende-se um elemento		
5	1	1	entende-se um grupo		
6	1	1	entende-se um material		
7	1	1	entende-se um máximo		
8	1	1	entende-se um reservatório		
9	1	1	entende-se um sistema		
10	1	1	entende-se um subconjunto		
11	1	1	entende-se um tubo		
12	1	1	entende-se um volante		

Esta configuração permite obter uma lista de ocorrências que confirma a nossa percepção inicial, de que existe um padrão. Nas 33 ocasiões em que [entende-se] se combina com a forma [um], em 20 dessas ocasiões encontra-se seguida da forma [dispositivo].

No exercício realizado no capítulo 5.1., sobre extração de termos, constatámos que, por vezes, uma unidade lexical obtida a partir da combinação de um termo com uma outra unidade lexical tinha grande probabilidade de ser um termo.

Verificámos também, que noutras ocasiões, quer por conhecimento insuficiente acerca do domínio, quer por falta de evidências extraídas a partir do contexto, não foi possível constituir uma unidade lexical como um possível termo.

No que se refere à situação em curso, apesar de termos estabelecido um padrão no que respeita a uma combinação específica de formas no *corpus*, não temos ainda matéria para delimitar uma unidade lexical.

Ao isolar as concordâncias relativas à combinação das formas representada por [entende-se um dispositivo], e a consultar o contexto através da função “File View”,

apercebemo-nos que, a forma [um] ocorre, nas 20 ocasiões, como um artigo definido de uma unidade frásica. Ao observar isoladamente as unidades frásicas constatamos que estas parecem pertencer a um contexto definatório.

Se de facto se trata de uma unidade frásica que corresponde a uma definição, implica que haja alguma coisa a ser definida, o que num texto de especialidade, terá grande probabilidade de corresponder a termo.

Com o objetivo de validar a nossa hipótese, procedemos à extração de alguns contextos, limitando a dimensão dos mesmos às formas compreendidas entre dois pontos, contendo o grupo constituído pelas formas [entende-se um dispositivo]:

1. “Por «válvula limitadora do enchimento a 80 %», **entende-se um dispositivo** que não permite encher o reservatório a mais de 80 % da sua capacidade.”
2. “Por «bomba de combustível», **entende-se um dispositivo** para fornecer GPL líquido ao motor, aumentando a pressão do reservatório com a pressão de alimentação da bomba de combustível.”
3. “Por «unidade de enchimento», **entende-se um dispositivo** que permite encher o reservatório; pode ser integrado na válvula limitadora do enchimento a 80 % ou localizar-se à distância, no exterior do veículo.”

Tomando como referência os exemplos atrás, é possível detetar um padrão. Em função da ocorrência desse padrão, consideramos que é correto afirmar que, em todos os exemplos, a preceder [entende-se um dispositivo], ocorre uma unidade lexical com forte probabilidade de ser um termo do domínio em estudo.

A partir da constatação anterior, consideramos que, no nosso *corpus*, a preceder [entende-se um], ocorre uma unidade lexical com forte probabilidade de ser constituída como um termo do domínio em estudo, apresentando para tal, os exemplos seguintes:

4. “Por «volante de substituição» (equipado com um módulo de almofada de ar), **entende-se um** volante destinado a modificar o veículo a motor e que, em relação ao volante original fornecido pelo fabricante, pode variar nas dimensões funcionais, na forma e/ou no material.”
5. “Por «sistema de almofada de ar de substituição», **entende-se um** sistema de almofada de ar destinado a modificar um veículo a motor e que, em relação ao sistema original de almofada de ar destinado pelo fabricante ao veículo em

causa, pode variar nas dimensões funcionais, na forma, no material ou no funcionamento.”

6. “Por «sensor», **entende-se um** componente concebido para detetar e transmitir ao controlador as condições de rotação da(s) roda(s) ou as condições dinâmicas do veículo.”

A observação do padrão que sobressai dos exemplos anteriores leva-nos a alargar a nossa hipótese em relação à forma [entende-se], em que passamos a estabelecer o seguinte: a preceder a forma [entende-se], seguida de um artigo, ocorre sempre uma unidade lexical com forte probabilidade de ser um termo do domínio em estudo, cujo formalismo pode ser assim representado: “Por  $x$  entende-se  $y$ ” em que  $x$  corresponde a um termo pertencente ao domínio e  $y$  corresponde a um artigo.

Vejamos nos exemplos seguintes se se confirma a nossa hipótese:

7. “Por «massa sem carga em ordem de marcha», **entende-se a** massa do veículo em ordem de marcha, sem ocupantes e sem carga, mas com combustível, fluido de arrefecimento, lubrificantes, ferramentas e uma roda de reserva (estas últimas, se fizerem parte do equipamento normalmente fornecido de origem pelo fabricante do veículo).”
8. “Por «feixe de fios», **entende-se o** conjunto dos condutores e cabos eléctricos que, no sistema completo de almofada de ar, ligam as várias partes umas às outras e eventualmente ao veículo.”
9. “Por «roda indiretamente controlada», **entende-se uma** roda cuja força de travagem é modulada a partir de informações provenientes do(s) sensor(es) de outra(s) roda(s) ( 1 ).”
10. “Por «tipo de um volante de substituição equipado com um módulo de almofada de ar», **entende-se os** volantes que não difiram entre si nos seguintes aspectos essenciais:”

No seguimento da análise anterior, ampliamos a escopo da nossa proposta. Verificamos que, nas ocasiões em que a seguir à forma [entende-se] ocorre um artigo (“a”; “o”; “uma”; “os”), este faz parte de uma unidade frásica maior. Assim, em relação ao formalismo anterior, “Por  $x +$  entende-se  $+y$ ” julgamos que a  $y$  corresponderá uma unidade frásica, podendo esta corresponder a um contexto definatório ou uma descrição e conter termos do domínio em estudo.

## 5.2.2. Análise da combinação das formas [entende-se por]

No seguimento da tarefa anterior, programamos o concordanceiro para extrair as concordâncias referentes à combinação da forma [entende-se] seguida da forma [por], à sua direita, obtendo os seguintes resultados:

**Figura 30- resultado da extração de concordâncias referente à combinação entre a forma [entende-se] com a forma [por]**

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Concordance Hits 30						
Hit	KWIC					
1	DE CRUZAMENTO (1) Para efeitos do presente anexo, <a href="#">entende-se por</a> : «acima», colocado por cima, segund					
2	. Para efeitos do disposto no presente anexo, <a href="#">entende-se por</a> : a) «Sistema completo», o lado					
3	DEFINIÇÕES Para efeitos do presente regulamento, <a href="#">entende-se por</a> : 2.1. «Roda», um elemento rotativo					
4	DEFINIÇÕES Para efeitos do presente regulamento, <a href="#">entende-se por</a> : 2.1. «Compatibilidade eletromagné					
5	DEFINIÇÕES Para efeitos do presente regulamento, <a href="#">entende-se por</a> : 2.1. «Homologação de um veículo»,					
6	. 2. DEFINIÇÕES Para efeitos do presente anexo, <a href="#">entende-se por</a> : 2.1. «Conceito de segurança», uma					
7	DEFINIÇÕES Para efeitos do presente regulamento, <a href="#">entende-se por</a> : 2.1. «Vidro temperado», uma vidra					
8	e uma vidraça de segurança. 9.2.1.1. Definições. <a href="#">Entende-se por</a> : 9.2.1.1.1. «Desvio ótico», o ângu					
9	ensaaios de comportamento da cabeça ao choque. <a href="#">Entende-se por</a> pequena janela, uma janela na					
10	ensaaios de comportamento da cabeça ao choque. <a href="#">Entende-se por</a> pequena janela, uma janela na					
11	Para efeitos do disposto no presente anexo, <a href="#">entende-se por</a> : 1.1. «Tipo de produto», todas as					
12	. 2. DEFINIÇÕES Para efeitos do presente anexo, <a href="#">entende-se por</a> : 2.1. «Veículo», qualquer veículo					
13	Para efeitos do disposto no presente regulamento, <a href="#">entende-se por</a> : 2.1. Sistema de retenção para cri					
14	lugar sentado para um adulto. Neste contexto, <a href="#">entende-se por</a> : 2.20.1. «Grupo de bancos de um					
15	DEFINIÇÕES Para efeitos do presente regulamento, <a href="#">entende-se por</a> : 2.1. «Homologação de um veículo»,					
16	essórios e ferramentas indicados pelo fabricante. <a href="#">Entende-se por</a> reservatório de combustível cheio o					
17	DEFINIÇÕES Para efeitos do presente regulamento, <a href="#">entende-se por</a> : 2.1. «Homologação de um banco», a					
18	efeitos da parte I do presente regulamento, <a href="#">entende-se por</a> : 2.1. «Sistema de alarme para veíc					
19	efeitos da parte II do presente regulamento, <a href="#">entende-se por</a> : 14.1. «Sistema(s) de alarme» (SA)					
20	efeitos da parte III do presente regulamento, <a href="#">entende-se por</a> : 26.1. «Imobilizador»: um disposit					
21	DEFINIÇÕES Para efeitos do presente regulamento, <a href="#">entende-se por</a> : 2.1. «Sistema de retenção para cr					
22	lugar sentado para um adulto. Neste contexto, <a href="#">entende-se por</a> : «Grupo de bancos de um					
23	DEFINIÇÕES Para efeitos do presente regulamento, <a href="#">entende-se por</a> : 2.1. «Tipo de pneu», os pneus					
24	.DEFINIÇÕES Para efeitos do presente regulamento, <a href="#">entende-se por</a> : 2.1.«Dispositivos para visão indir					
25	.DEFINIÇÕES Para efeitos do presente regulamento, <a href="#">entende-se por</a> : 12.1.«Pontos oculares do condutor»					
26	disposto na parte II do presente regulamento, <a href="#">entende-se por</a> : 14.1.1. «Homologação de um veículo					
27	DEFINIÇÕES Para efeitos do presente regulamento, <a href="#">entende-se por</a> : 2.1.«Homologação de um veículo», a					

Voltamos a usar o mesmo processo aplicado no exercício anterior para a combinação de formas [entende-se um], na tentativa de isolar um padrão recorrente presente no *corpus* com base na combinação das formas [entende-se] e [por].

Procedemos à personalização das ferramentas do concordanceiro para isolar as combinações possíveis mais frequentes de [entende-se por] com as formas que ocorrem à sua direita, obtendo os seguintes resultados:



**Figura 31 - lista das formas mais frequentes associadas à direita da forma [entende-se por]**

Concordance				Concordance Plot	File View	Clusters/N-Grams	Collocates	Word Lis
Total No. of Cluster Types				26				Total No. of Cluster Tokens
								30
Rank	Freq	Range	Cluster					
1	3	3	entende-se por: 2.1. «homologação					
2	2	1	entende-se por pequena					
3	2	2	entende-se por: 2.1. «sistema					
4	1	1	entende-se por reservatório					
5	1	1	entende-se por «para					
6	1	1	entende-se por: 1.1. «tipo					
7	1	1	entende-se por: 14.1. «sistema					
8	1	1	entende-se por: 2.1. sistema					
9	1	1	entende-se por: 2.1. «compatibilidade					
10	1	1	entende-se por: 2.1. «conceito					
11	1	1	entende-se por: 2.1. «veículo					
12	1	1	entende-se por: 2.1. «vidro					
13	1	1	entende-se por: 2.20.1. «grupo					
14	1	1	entende-se por: 26.1. «imobilizador					
15	1	1	entende-se por: 9.2.1.1.1. «desvio					
16	1	1	entende-se por: «grupo					
17	1	1	entende-se por: 12.1.«pontos					
18	1	1	entende-se por: 14.1.1. «homologação					
19	1	1	entende-se por: 2.1. «conceito					
20	1	1	entende-se por: 2.1. «modo					
21	1	1	entende-se por: 2.1. «roda					
22	1	1	entende-se por: 2.1. «tipo					
23	1	1	entende-se por: 2.1.«dispositivos					
24	1	1	entende-se por: 2.1.«homologação					
25	1	1	entende-se por: a					
26	1	1	entende-se por: «acima					

Constatamos que não há uma forma que demonstre uma elevada frequência, mas verifica-se uma grande variedade de formas diferentes, com a sua frequência de ocorrência a variar entre 1 e 3. As formas atrás mencionadas parecem fazer parte de um conjunto de formas agrupadas por forma a constituir uma unidade frásica, destacando-se de seguida alguns exemplos:

1. “**Entende-se por** pequena janela, uma janela na qual seja impossível inscrever um círculo com um diâmetro de 150 mm.”
2. “**Entende-se por** reservatório de combustível cheio o enchimento de pelo menos 90 por cento da sua capacidade.”
3. “Para efeitos do presente regulamento, **entende-se por**: 2.1. «Homologação de um banco», a homologação de um tipo de banco como componente, no contexto da proteção dos ocupantes dos bancos voltados para a frente, no que diz respeito à resistência destes e à conceção dos seus encostos;”

Ao analisar os exemplos atrás, podemos afirmar que a unidade frásica resultante da combinação das formas estabelece um contexto definatório, contendo unidades lexicais com forte probabilidade de serem termos pertencentes ao domínio em estudo.

Em semelhança às nossas constatações no capítulo anterior, propomos o formalismo “entende-se por  $z$ ”, com o intuito de dar conta das possibilidades de extrair unidades lexicais candidatas a termos do domínio. Como tal, a  $z$  corresponderá uma unidade frásica que corresponde a um contexto de tipo definatório, podendo este conter unidades lexicais com possibilidade de corresponderem a termos do domínio em estudo.

### **5.2.3. Observações:**

Como já referimos anteriormente, o objetivo deve ser apurar ao máximo os critérios de pesquisa, na tentativa de isolar padrões de frequência que possibilitem apurar resultados com vista à seleção de unidades lexicais candidatas a termos.

Nesta situação em particular, demonstrámos como apurar candidatos a termos, a partir de uma forma pré-selecionada. Pudemos constatar que ao depurar as formas que ocorrem à direita da forma selecionada, foi possível verificar a ocorrência de combinações específicas que se traduziram na obtenção de unidades lexicais com forte probabilidade de se constituírem como termos do domínio em estudo.

Reconhecemos que as nossas intuições não são suficientes para afirmar que chegámos a extrair termos pertencentes ao domínio, mas estabelecemos uma metodologia que permite produzir uma lista de possibilidades, lista essa que pode ser apresentada a um especialista, cuja experiência como tal, permitirá obter as confirmações necessárias.

Para além de candidatos a termos pudemos ainda, através da experimentação da nossa metodologia, isolar contextos específicos, com carácter definatório, sendo estes um tipo de contexto importante para a construção de terminologias, por neles se encontrar o termo e a informação necessária à sua definição.

Assim podemos afirmar que os processos aplicados no capítulo 5.2.1. permitiram obter candidatos a termos e isolar contextos de tipo definatório, enquanto em 5.2.2., só foi possível isolar os contextos de tipo definatório, como se verifica nos formalismos propostos.

## Conclusão

A elaboração deste trabalho teve como meta principal criar, descrever e sistematizar uma metodologia de análise de *corpora* e extração de termos de especialidade, utilizando para tal, recursos disponíveis na internet.

Considerando a especificidade da construção de *corpora* de especialidade, começámos por tentar estabelecer a dimensão do domínio da Segurança Automóvel, recorrendo à consulta de dicionários online de língua corrente, complementando a informação obtida nestes, com uma pesquisa mais abrangente recorrendo a um motor de busca da internet.

De seguida escolhemos uma fonte para os textos de especialidade, sendo que para tal, baseámo-nos em critérios de fiabilidade implicitamente garantidos pelo estatuto atribuído à entidade repositora dos textos.

A análise do *corpus* foi realizada com recurso a uma ferramenta de extração de concordâncias, o AntConc, possibilitando destacar unidades lexicais, retiradas a partir de contextos, com probabilidade de serem termos pertencentes ao domínio da Segurança Automóvel.

Consideramos ter cumprido o objetivo principal, o de descrever os processos a seguir quando se pretende trabalhar com *corpora* de especialidade. No entanto, tínhamos como meta a atingir, três objetivos de cariz distinto, mas relacionados entre si de forma sucessiva.

Na nossa opinião, uma metodologia não se pode afirmar sem antes ser aplicada, por forma a obter a confirmação acerca dos processos estabelecidos. Consideramos ter definido um número significativo de processos metodológicos, mas sabemos que a utilidade desta metodologia está dependente das características dos seus futuros utilizadores.

Julgamos que seria benéfico testar a metodologia em ambiente prático de formação. O ambiente de formação de línguas de especialidade tem dois tipos de beneficiários da metodologia de extração e análise de *corpora*, a saber, formadores e formandos.

A distinção atrás feita é importante, pois a cada um destes utilizadores corresponderão necessidades diferentes de análise, sendo que a nossa metodologia poderá ser adaptada de forma diferente, consoante as suas necessidades.

No que respeita aos formadores, a metodologia poderá ser aplicada em função de dois objetivos: em primeiro a necessidade do formador construir *corpora* para ser analisado pelos

formandos; em segundo a necessidade de analisar *corpora* construído pelos formandos com o objetivo de identificar as dificuldades destes na compreensão das matérias lecionadas.

No que respeita aos formandos, a sua metodologia estará sempre condicionada pela metodologia de ensino adotada pelo formador. Lembre-se, contudo, que a análise de *corpora* pode ser efetuada de forma autónoma, o que significa que os formandos têm sempre a possibilidade de, por eles próprios, adotar a metodologia e realizar a sua aplicação por forma a cumprir objetivos pessoais na procura e análise de informação contida em *corpora*.

O nosso objetivo final é, após os anteriores estarem completados, construir um recurso terminológico, no caso uma base dados textual, referente ao domínio da Segurança automóvel.

Consideramos que este recurso é de grande utilidade, pois cada vez mais linguistas, mas não só, recorrem à análise de *corpora* para obter dados obtidos a partir do uso real da língua, sobretudo em contextos de especialidade.

Apesar de ser possível obter textos na internet e juntá-los num *corpus*, a existência de recursos criados a partir de metodologia atestada, têm um grande valor, pois permitem poupar tempo e trabalho.

## Bibliografia

- Aijmer, K. (Ed.). (2009). *Corpora and language teaching* (Vol. 33). John Benjamins Publishing
- Antia, B. (2000). *Terminology and language planning: An alternative framework of practice and discourse* (Vol. 2). John Benjamins Publishing.
- Barnbrook, G., Danielsson, P., & Mahlberg, M. (Eds.). (2005). *Meaningful texts: the extraction of semantic information from monolingual and multilingual corpora*. A&C Black.
- Bhatia, V., Hernández, P. S., & Pérez-Paredes, P. (Eds.). (2011). *Researching specialized languages* (Vol. 47). John Benjamins Publishing.
- Bondi, M., & Scott, M. (Eds.). (2010). *Keyness in texts* (Vol. 41). John Benjamins Publishing.
- Bourigault, D., Jacquemin, C., & L'Homme, M. C. (Eds.). (2001). *Recent advances in computational terminology* (Vol. 2). John Benjamins Publishing.
- Bowker, L., & Pearson, J. (2002). *Working with specialized language: a practical guide to using corpora*. Routledge.
- Budin, G. and Wright, S-H. (1997) - Handbook of Terminology Management Vol. 1. Basic Aspects of Terminology Management. Amsterdam: Benjamins.
- Cabré, M.T. (1993). *La Terminología: teoría, metodología, aplicaciones*. Traducción castellana de Carles Tebé. Barcelona: Editorial Antártida
- Cabré, M.T. (1998) – A propos de la notion de qualite en Terminologie. *La Banque des Mots*, 7-34
- Cabré Castellví, M. T. (2003). Theories of terminology: Their description, prescription and explanation. *Terminology*, 9(2), 163-199.
- Costa, R. (2001) - Pressupostos teóricos e metodológicos para a extracção automática de unidades terminológicas multilexémicas. Tese de doutoramento em Terminologia. Universidade Nova de Lisboa

- Costa, R. (2006). Texte, terme et contexte. Septièmes Journées Scientifiques AUF-LTT «Mots, termes et contextes», Bruxelles, 8.
- Correia, M. (1998). Neologia e terminologia. Terminologia: questões teóricas, métodos e projectos, 59-74.
- Cruse, D. A. (1986). *Lexical semantics*. Cambridge University Press.
- Cruse, D. A. (2006). *A glossary of semantics and pragmatics*. Edinburgh University Press.
- Davies, A., & Elder, C. (Eds.). (2008). *Handbook of applied linguistics*. John Wiley & Sons.
- Depecker, L. (2002). *Entre signe et concept: éléments de terminologie générale*. Presses Sorbonne Nouvelle.
- Estopà, R. (2009). Los extractores de terminología: logros y escollos. In Terminología y Sociedad del conocimiento (pp. 117-146).
- Felber, H. (1984). Terminology Manual.
- Gavioli, L. (2005). *Exploring corpora for ESP learning* (Vol. 21). John Benjamins Publishing.
- Geeraerts, D. (2010). *Theories of lexical semantics*. Oxford University Press.
- Ghadessy, M., Henry, A., & Roseberry, R. L. (Eds.). (2001). *Small corpus studies and ELT: theory and practice* (Vol. 5). John Benjamins Publishing.
- Granger, S., Hung, J., & Petch-Tyson, S. (Eds.). (2002). Computer learner corpora, second language acquisition, and foreign language teaching (Vol. 6). John Benjamins Publishing.
- Halliday, M. A. K. (2004). Lexicology. *Lexicology and Corpus Linguistics*, 1-22.
- Habert, B., Nazarenko, A., & Salem, A. (1997). *Les linguistiques de corpus*. Colin.
- Hoey, M. (2007). *Text, discourse and corpora: Theory and analysis*. A&C Black.
- ISO 1087-1:2000. Terminology Work – Vocabulary – Part 1: Theory and Application. Geneva: ISO.
- Kageura, K. (2002). *The Dynamics of Terminology: A Descriptive Theory of Term*

- Formation and Terminological Growth*. Amsterdam/Philadelphia, Editora John Benjamins Publishing.
- Kennedy, G. (1998). *An introduction to corpus linguistics*. Routledge.
- L'Homme, M. C. (2004). *La terminologie: principes et techniques*. Pum.
- Lerat, P. (1995). *Les langues spécialisées*. Presses universitaires de France.
- Lüdeling, A. (Ed.). (2008). *Corpus linguistics* (Vol. 1). Walter de Gruyter.
- Meyer, C. F. (Ed.). (2002). *English corpus linguistics: An introduction*. Cambridge University Press.
- Nesselhauf, N. (2005). *Collocations in a learner corpus* (Vol. 14). John Benjamins Publishing.
- Nesselhauf, N. (2011) *Corpus Linguistics: A Practical Introduction*
- Pearson, J. (1998). *Terms in context* (Vol. 1). John Benjamins Publishing.
- Rey, A. (1979). *La terminologie: noms et notions*. Paris: Presses universitaires de France
- Sager J. C., (1990) *Practical Course in Terminology Processing*. John Benjamins Publishing.
- Sager, J. C. (Ed.). (2000). *Essays on definition* (Vol. 4). John Benjamins Publishing.
- Saint-Dizier, P., & Viegas, E. (1995). *Computational lexical semantics*. Cambridge University Press.
- Santos, C. (2010). *Terminologia e ontologias: metodologias para representação do conhecimento*. Tese de Doutorado. Universidade de Aveiro
- Sardinha, T. B. (2000). *Linguística de corpus: histórico e problemática*. *Delta*, 16(2), 323-367.
- Sardinha, T. B. (2006). *Pesquisa em linguística de corpus com Wordsmith tools*. Campinas: Mercado de Letras.
- Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education* (Vol. 22). John Benjamins Publishing.

- Silva, R., Costa, R., & Ferreira, F. (2004). Entre langue générale et langue de spécialité une question de collocations. *Ela. Études de linguistique appliquée*, (3), 347-359.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Sinclair, J. (2003). *Reading concordances: an introduction*. Pearson Longman
- Sinclair, J. M. (Ed.). (2004). *How to use corpora in language teaching* (Vol. 12). John Benjamins Publishing.
- Sinclair, John (2005) *Corpus and Text - Basic Principles*. In Martin Wynne, ed. *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxbow Books; 1-16.
- Stubbs, M. (1996). *Text and corpus analysis: Computer-assisted studies of language and culture*. Oxford: Blackwell.
- Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell Publishers.
- Teubert, W. (1999). *Corpus Linguistics. A Partisan View*. *TELRI Newsletter*, 8(99), 4-19.
- Teubert, W. (2005). My version of *corpus linguistics*. *International journal of corpus linguistics*, 10(1), 1-13.
- Teubert, W. (Ed.). (2007). *Text corpora and multilingual lexicography* (Vol. 8). John Benjamins Publishing
- Université de Paris 13. Centre d'études lexicales. Colloque (1988: Paris), & Chaurand, J. (1990). La définition/[actes du Colloque " la Définition", organisé par le CELEX (Centre d'études du lexique) de l'Université de Paris-Nord (Paris 13, Villetaneuse) à Paris, les 18 et 19 novembre 1988];[éd. par Jacques Chaurand]. Larousse.
- Van Sterkenburg, P. (Ed.). (2003). *A practical guide to lexicography* (Vol. 6). John Benjamins Publishing.
- Viana, V., Zyngier, S., & Barnbrook, G. (Eds.). (2011). *Perspectives on corpus linguistics* (Vol. 48). John Benjamins Publishing.



- Wichmann, A., & Fligelstone, S. (1997). *Teaching and language corpora*. London, New York: Longman
- Williams, G., & de Corpus, J. D. L. L. (2005). *La linguistique de corpus*.
- Wüster, E. (2010). *Introducción a la teoría general de la terminología ya la lexicografía terminológica*. Documenta Universitaria. Empúries

## Índice De Figuras

Figura 1 - Dicionário Priberam .....	4
Figura 2 - Dicionário Infopédia .....	4
Figura 3 - Resultado da pesquisa com motor de busca .....	6
Figura 4 - Estrutura do domínio .....	12
Figura 5 – Classificação europeia de veículos .....	15
Figura 6 – Menu de trabalho da base de dados EUR-lex .....	24
Figura 7 – Ambiente de Trabalho do AntConc .....	27
Figura 8 - Menu principal .....	28
Figura 9 - File .....	28
Figura 10 – Global Settings .....	29
Figura 11 – Tool Preferences .....	29
Figura 12 - Menu de ferramentas .....	30
Figura 13 - Menu de busca .....	31
Figura 14 – Lista de formas ordenada pelas frequências mais altas .....	31
Figura 15 – Lista de palavras ordenada alfabeticamente .....	32
Figura 16 – Identificação dos ficheiros de texto .....	33
Figura 17 – Extrato da lista de frequências do <i>corpus</i> do domínio da Segurança Automóvel .....	37
Figura 18 - Resultado da extração das concordâncias do termo ‘segurança’ ....	40
Figura 19 – Identificação dos textos que contêm o termo ‘segurança’ .....	40
Figura 20- lista das dez unidades lexicais mais frequentes associadas à esquerda do termo ‘segurança’ .....	41
Figura 21 - lista das unidades lexicais criadas a partir do termo ‘cinto de segurança’	

.....	42
Figura 22 - lista das dez unidades lexicais mais frequentes associadas à direita do termo 'segurança' .....	44
Figura 23 – Resultado das concordâncias das formas [para adultos] precedidas de 'cinto de segurança' .....	45
Figura 24 – Resultado das concordâncias das formas [para adultos] precedidas de 'cintos de segurança' .....	45
Figura 25 - Resultado das concordâncias da unidade lexical /segurança de funcionamento/ .....	46
Figura 26 - Resultado da extração das concordâncias da forma [entende-se] .....	49
Figura 27 – lista das dez unidades lexicais mais frequentes associadas à direita da forma [entende-se].....	50
Figura 28 - Resultado da extração de concordâncias referente à combinação entre a forma [entende-se] com a forma [um] .....	51
Figura 29 - lista das dez unidades lexicais mais frequentes associadas à direita da forma [entende-se um].....	52
Figura 30 - Resultado da extração de concordâncias referente à combinação entre a forma [entende-se] com a forma [por] .....	55
Figura 31 - lista das formas mais frequentes associadas à direita da forma [entende-se por] .....	56

## **Índice De Tabelas**

Tabela 1 – Definições para ‘automóvel’ e ‘segurança’ .....	5
Tabela 2 – Áreas de atividade da responsabilidade da CE, IMT E ANSR.....	12
Tabela 3 – Opções do Menu de ferramentas .....	30
Tabela 4 – Classificação gramatical formas ocorrentes à direita de [entende-se] .....	50